

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 29, 2002, 14:09:17 ; Search time 70.54 Seconds

(without alignments)  
213.816 Million cell updates/sec

Title: US-09-548-971B-2

Perfect score: 1055

Sequence: 1 MNGMYKKKGVCDSCVSSKS.....QIGAPMANPSYLCYHNSQP 198

Scoring table:

BLOSUM62

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1055	100.0	198	2	T01340	hypothetical prote
2	397.5	37.7	224	2	T49923	hypothetical prote
3	360	34.1	231	2	T45579	hypothetical prote
4	191	18.1	912	2	T48153	hypothetical prote
5	178	16.9	298	2	C96690	unknown protein F2
6	175	16.6	279	2	B86402	protein T22C5.19 f
7	170.5	16.2	349	2	T05216	hypothetical prote
8	169.5	16.1	367	2	G85362	hypothetical prote
9	157.5	14.9	350	2	E84634	hypothetical prote
10	156.5	14.8	310	2	T01090	hypothetical prote
11	149	14.1	297	2	B86161	hypothetical prote
12	143.5	13.6	524	2	H86228	hypothetical prote
13	137	13.0	264	2	B86432	protein T518.12 f1
14	136.5	12.9	315	2	T49982	hypothetical prote
15	136	12.9	266	2	T00464	hypothetical prote
16	134	12.7	415	2	B85436	hypothetical prote
17	128	12.1	467	2	G84545	bHLH transcription
18	126.5	12.0	490	2	A84586	hypothetical prote
19	122.5	11.6	250	2	G96765	hypothetical prote
20	122.5	11.6	423	2	H84860	hypothetical prote
21	121.5	11.5	314	2	T06032	hypothetical prote
22	120.5	11.4	296	2	A85065	hypothetical prote
23	118.5	11.2	300	2	D85431	hypothetical prote
24	118	11.2	285	2	C86423	unknown protein, 1
25	118	11.2	442	2	T47788	hypothetical prote
26	116.5	11.0	262	2	B85093	hypothetical prote
27	116.5	11.0	380	2	T00937	hypothetical prote
28	114.5	10.9	259	2	F86475	hypothetical prote
29	114.5	10.9	347	2	T06329	synbiotic ammonium

30	113.5	10.8	126	2	S65807	transcription acti
31	113.5	10.8	596	2	D96661	probable transcrip
32	113	10.7	329	2	T01333	hypothetical prote
33	113	10.7	644	2	A42220	helix-loop-helix p
34	112.5	10.7	562	2	S16594	regulatory protein
35	112	10.6	121	2	S72496	transcription acti
36	112	10.6	129	2	S65803	transcription acti
37	112	10.6	586	2	S65802	transcription acti
38	111	10.5	146	2	T14681	myc-like regulator
39	111	10.5	379	2	H96548	unknown protein f1
40	111	10.5	662	2	G84903	probable bHLH tran
41	110.5	10.5	318	2	T52428	helix-loop-helix p
42	110.5	10.5	324	2	T05273	hypothetical prote
43	110.5	10.5	348	2	B96620	protein T30E16.21
44	110	10.4	121	2	S65809	transcription acti
45	110	10.4	288	2	D84616	probable bHLH tran

## ALIGNMENTS

RESULT  
T01340  
hypothetical protein F6N15.18 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 14-May-1999  
C:Accession: T01340  
R/Ryan, E.; Edwards, J.; Pape, K.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of A. thaliana F6N15.  
A:Reference number: Z14297  
A:Accession: T01340  
A>Status: translated from GH/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-198 <RYA>  
A:Cross-references: EMBL:AF069299; NID:g3193311; PID:g3193315  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Note: F6N15.18

Query Match 100.0%; Score 1055; DB 2; Length 198;  
Best Local Similarity 100.0%; Pred. No. 7e-83;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNGMYKKKGVCDSCVSSKSRSNSPKRSMPEQPHLLMDWNKANDLTQEHAAFLNDP 60  
|||||  
Db 1 MNGMYKKKGVCDSCVSSKSRSNSPKRSMPEQPHLLMDWNKANDLTQEHAAFLNDP 60  
QY 61 HHLMDPPPELILHDEDEYEDMDAMKEMQYMAVQPVDPIDPATVPKPNRRVRIISD 120  
|||||  
Db 61 HHLMDPPPELILHDEDEYEDMDAMKEMQYMAVQPVDPIDPATVPKPNRRVRIISD 120  
QY 121 DPQTVARRRRERISEKIRILKRIYVPGAKMDTASMLDEAIRYTKFLKROVRILOPHSOI 150  
|||||  
Db 121 DPQTVARRRRERISEKIRILKRIYVPGAKMDTASMLDEAIRYTKFLKROVRILOPHSOI 180  
QY 181 GAPMANPSYLCYHNSQP 198  
|||||  
Db 181 GAPMANPSYLCYHNSQP 198  
RESULT 2  
T49923  
hypothetical protein F17I14.60 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49923  
R/Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-224 <BEV>  
A:Cross-references: EMBL:AL353994; GSPDB:GN00063; ATSP:F17114.60  
A:Experimental source: cultivar Columbia; BAC clone F17114  
C:Genetics:  
A:Gene: ATSP:F17114.60  
A:Map position: 5

Query Match 37.7%; Score 397.5; DB 2; Length 224;  
Best Local Similarity 46.6%; Pred. No. 9,1e-27;  
Matches 88; Conservative 28; Mismatches 46; Indels 27; Gaps 3;

QY 22 SNHSPKSMMEPQPHLLMDWNKANDLLTQEHAAFLNDPHTLMD----- 65  
DB 2 NNYMNPSTLQNTYTNWNIINSNNKNDHHOHNDPIGMANDQYTOHLIFNDESSSR 61  
QY 67 -P-PE-----PETLIHDEDEEYD-----EDMDAMKEMQYMAVQPVDPATVPKPNKR 114  
DB 62 FPLSSSLTTTTLTSGDQDEDEDEEPLLELGAMKEMMYKIAMQSVDPATVPKPKR 121  
QY 115 NVRISSDPQTVVARRRRERISEKIRILKRIYVGGAKMDTASMLDEAIRYTKFLKQVRIL 174  
DB 122 NVRISSDPQSVAAARRRRERISERIRILOPLVPGTKMDTASMLDEAIRYTKFLKQIRLL 181  
QY 175 QPHSCIGAP 183  
DB 182 NNNTGYTPP 190

RESULT 3  
T45579  
hypothetical protein F11C1.170 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45579  
R:Bargues, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.; May  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: 223007  
A:Accession: T45579  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-231 <BAR>  
A:Cross-references: EMBL:AL132976  
A:Experimental source: cultivar Columbia; BAC clone F11C1  
C:Genetics:  
A:Map position: 3  
A:Note: F11C1.170

Query Match 34.1%; Score 360; DB 2; Length 231;  
Best Local Similarity 45.8%; Pred. No. 1.5e-23;  
Matches 82; Conservative 26; Mismatches 37; Indels 34; Gaps 6;

QY 22 SNHSPKSMMEPQPHLLMDWNKANDLLTQEHAAFLNDPHTLMDPPETLIHLEDEE 80  
DB 23 SNSNP-----NPNPHNIM-LSESN---IHPFFNPTSHSLPFD--QTPHHPGLN 68  
QY 81 Y-----DEMDAMKEMQYMAVQPVDPATVPKPNRRNVRISSD 121  
DB 69 FRYAPSSSSLPKRGCCSDNANMAAMREMIFRIAVQPIHIDPESVKKPKRKNVRISKD 128  
QY 122 PQTVAARRRRERISEKIRILKRIYVGGAKMDTASMLDEAIRYTKFLKQVRILQPHSQI 180  
DB 129 PQSVAAARRRRERISERIRILOPLVPGTKMDTASMLDEAIRYTKFLKQVOSLEEHAVV 187

RESULT 4  
T48153  
hypothetical protein T1008.20 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48153  
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Lemcke, K.; Maye  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: 224486  
A:Accession: T48153  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-912 <BEV>  
A:Cross-references: EMBL:AL161746  
A:Experimental source: cultivar Columbia; BAC clone T1008  
C:Genetics:  
A:Map position: 5  
A:Introns: 168/1; 260/3; 551/2; 750/3; 807/3  
A:Note: T1008.20

Query Match 18.1%; Score 191; DB 2; Length 912;  
Best Local Similarity 61.3%; Pred. No. 2e-08;  
Matches 38; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 110 KPNRRNVRISSDPQTVVARRRRERISEKIRILKRIYVGGAKMDTASMLDEAIRYTKFLK 169  
DB 35 KRSRSTSTLSTPQSVAAARRRRRISDRFKILQSVVGGAKMDTVSMLEAISYVKFLKA 94  
QY 170 QV 171  
DB 95 QI 96

RESULT 5  
C96690  
unknown protein F28G11.9 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96690  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alor  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talli  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C96690  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-298 <STO>  
A:Cross-references: GB:AE005173; NID:g11054535; PIDN:AAG27834.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F28G11.9  
A:Map position: 1

Query Match 16.9%; Score 178; DB 2; Length 298;  
Best Local Similarity 25.9%; Pred. No. 6.6e-08;  
Matches 55; Conservative 30; Mismatches 69; Indels 58; Gaps 6;

QY 17 SSKRSNHSFKRSM---EPQPH-LMDWNKANDLLTQEHAAFLNDPHTLMDPPETLIHLEDEE 56  
DB 50 SSTMNSDHOHQGFYFVPSGEDHHNSLMDFNSSFLNFDHHSFPRAISGGSSGGCGF 109  
QY 57 -----LNDPHTLMDPP-----PETLIHLEDEEYDEMDA 87  
DB 110 SFLEGNMMSYFTNMNHQHMDIISPRSTETPGQKDWLYSDSTVVTGSRNESLSPKSA 169  
QY 88 MKEMQYMAVQPV-----VDIDPATVPKPNRRNVRISSDPQTVVARRRRERISEKIRILK 142  
DB 170 GKRSHTGESTQPSKLISSGVYTKTKPKPTTS----PKDPSLAAKNRERISERILKIQ 225  
QY 143 RIVPGAKMDTASMLDEAIRYTKFLKQVRIL 174

gene

TGLSPSSDVLVCLPVRNI VSKPKIIPHLVLKITSLLLKVVVNODSVLKFYI  
VFILMNTCLYIFVLGASLIIPYVLPFLICGSGHVIYEPIIMLVVEPCDLICSI  
QSPKGPDCITVDQNGNSTEMVGEDSAANSIPFNTGNLAIAKSVRLDISFKPSH  
TGLQTQLVKDITEQFPATPLVLVKQFLADRLDQSYSGLSYCLVRCILLLK"  
complement(23534..25288)  
/gene="F6N15.15"  
complement(join(23534..23629,23711..23845,23934..23999,  
24049..24092,24457..24490,24753..24823,25132..25288))  
/gene="F6N15.15"  
/note="contains similarity to Lotus japonicus RING-finger  
protein (GB:236750)"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAC19316.1"  
/db\_xref="GI:3193334"

CDS

/translation="MSLSEENELEFDLNEIPEKTVSDGDGDFLFDLKIIPREETV  
NSSDEDFADHLHRNQATKRRKKLRPKVFRHIEKVISFSFYLVLRASSFLFDLNM  
GLRMDTPYQEI RMDTDHMTYEDLIQICNNMGYENGSKASNIDKCLRNTPSEFOSI.  
ADKICCIDGDFQKRAGVGLNCGHNFHVNCVKPWILTKK"  
26808..27437  
/gene="F6N15.9"  
26808..27437  
/gene="F6N15.9"  
/note="contains similarity to pectinesterases"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAC19307.1"  
/db\_xref="GI:3193325"

gene

/translation="MAANKLFEVLLSLPLIIFSATATSSKDYDTKAVVHSMCRTIL  
YPKLCVRSMSRYVRSRAVONPRDLARFALKASLYRAKYTKALKEVKNSETTLRQY  
YASVHDCLDQIRDSVNOLSLIAEIDRSVRSGKSGDLWHIINLQTMSTALIDAE  
TCVSQFPGRRMSKLKATIKGKRVNEETTSNALAFIEHYAAARYRARRP"  
27662..27734  
/note="codon recognized: AAG; Lys"  
/product="tRNA-OTHER"  
28096..30456

tRNA

gene

Query Match 59.68; Score 3842; DB 8; Length 95643;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	8	ccatctactatccggtgtgtgaaccccttaagcttttgaagactactagaataatgcaaat	67
Db	39816	CCATCTACTATCCGGTGTGTGACCCCTTAAAGCCTTTTGAAGACTACTAGAAATATGCAAAAT	39757
QY	68	accataatgtccatatccatcccttcttltgttgaactgaacatctcaatttgyaaaa	127
Db	39756	ACCATAATGTCCATAICCATCCCTTTCTTTGTGTGAACCTGAACATTTCTAATTTGTAAAA	39697
QY	128	gaaaaaaccttatgttaataacccgtgaagcaaaaaataatctcatcataatattt	187
Db	39696	GAAAAAACCTTATGTTAATATCACCCGTAGGCAAAAAAATAATCTCATCATATTTAAATTTT	39637
QY	188	tattataagattatacatctctctcgttgaagagtactccaattgcaagtgtgtatta	247
Db	39636	TATTATAGCATTATACATTCTCTCGTTGTAAGAGTTACTCCAAATGCCAAGTGTGTATTA	39577
QY	248	actaataaaaaagagcaaaatgaagcttataatlaattgatgtgcatagtactgta	307
Db	39576	ACTAATAAAAAAGCAGCAAGTAGGAAGCTTATATAATTGAATGATGTGCATAGTACTGGTA	39517
QY	308	tactgtgatgaataataacaagtatgaacattaatgtcatgtaaacgggtatttgtcttg	367
Db	39516	TATGTGTGATGAATATAACAGTATGAACATTAAATGATGTGCATAGTACTGGTA	39457
QY	368	aactcatlaaaggaatgtgaaaaagaagtgtgaggtctcatcttgaanaatttaccctct	427
Db	39456	AACTCATTAAGGCAATGTGAANAAGAGATGTGAGGTCTCATYTTGAAAAATTATCTTCT	39397
QY	428	agctttgtcagatcttaaacctatgaaatgaaagcaacatataaagaatttcatgtgacaa	487
Db	39396	AGCTTTGTGCAATTTTAAATCTATGAATGAACGCAACATATAGAAAATTTCATGTGACAA	39337

QY	488	cgacaattagaacggtactcttaattagaccgaatttaagtaataactatataataat	547
Db	39336	CGACATTTAGACGGGTATCTTAATTAGACCCGATTAATTAGTAATATACTTATATATTAAT	39277
QY	548	tagtgtgattataaagtttaactatccacttgagaatttaacaatgagcaataccitaa	607
Db	39276	TAGTGTGATTATAAGTTTACTTATCCACTTGAGAAATTTAAACAATGGCAATACCTTAA	39217
QY	608	tgtcgaagaagaccggtcccccacttcgtgtaatgaqtatgagggaagatccgtttaat	667
Db	39216	TGTGAAAGAAGCCGTCCTCCACTTCGTGTAATGAGTTATGGGGAGAGATCCTGTTAAAT	39157
QY	668	cgtcaaatataacaacttaagaactagaanaattgacacccaataatcaagaacgttg	727
Db	39156	CGTCAATTAACAACCTTAAGAACTAGAAATTTGACACCAAAAATCATAAAGAACGTTG	39097
QY	728	aagaagtcattatcgtatccagctcatatltccctagctagaatcaaatcaagccgttg	787
Db	39096	AAGAAGTCATTTATCGTATCCAGCTCATATATTCTTAGCTAAGATCAAAATCAAGGCCGTTG	39037
QY	788	aaagggcttgaagaaaatgtcgaagaaacccgtggggttagaagaaaagacaagaataag	847
Db	39036	AAAGGCTGTGAAGAAATGTGGAAGAAACCGTGGGTTTAGAACAAGACAAGAAATAG	38977
QY	848	aagaacaatgatgtttaaatttgcctatttctgtgtatagaggtgtcacaagaagagag	907
Db	38976	AAGAACAATGATGTTAAATTGCCATTTTGGTGTATAGAGTTGTCAAAAAGAGAGAGAG	38917
QY	908	agaagaataattagtcacaataatagagcaccataaaaatggaagacatgtgtgaagtaactat	967
Db	38916	AGAAGAAATTTAGTCAAAATTAATGAGCAGCTAAAAATGGAGACATGTCTTGAGTAACATAT	38857
QY	968	tacaagagcgactatgcttcccttatgagcaatgatataccaacccaagtgcacgcctcct	1027
Db	38856	TACAAGAGCGACTTATGCTTCCTTATGGCAATGATATCCAAACCAAGTGCAACGCTCCT	38797
QY	1028	tcttgcctaatcttcgtaaaagtctcctccctctccttgccttaggaaaaaacctagaana	1087
Db	38796	TTTTTGCCCTAATTTCGTAAGTCTCTCTCCTTCTGCTTAGGAAAAACCTTAGAAA	38737
QY	1088	tttaatcccttgttctgtacttgcctttttagaataaccatgatcttgaccacacactatt	1147
Db	38736	TTTAATCCCTTGTCTGTGATCTTGCTTTTGAGTAACCATGATTTTGACCAACACACTAT	38677
QY	1148	tcttcatccttctgtgtcctataagatttgccttataatgtgttcttctgtattgcctcg	1207
Db	38676	TCTTCTATCTTTTGTGTGTATAGGATTTTGCTTTATATGTGTCTTGTATTTGCTCCGT	38617
QY	1208	acgtacgtatacgaatttcaaatggtctatacaacaaggttataataactagcacaaatgagt	1267
Db	38616	ACGTACGTATACGAATTTAAATGGTTATTAACAAGCTTATATTAACCTAGCACAAATGAGT	38557
QY	1268	ccatgaaatttgttagcgaaaaaagtagaataatatlgagctctttaaacygcaatatata	1327
Db	38556	CCATGAAATTGTTAGCGCAAAAAGGTAGAAATATATIGAGCTTTTAAACGGCAATATATA	38497
QY	1328	taatttgcctgcacaaactagctttaatcatgatctaataatgataltttcttcaattccctt	1387
Db	38496	TAATTTGCTGCAAAACTTAGCTTTAATCATGATCTAATGATATTTCTTTAATTCTCT	38437
QY	1388	tgccaataatcatcacatgcacggatttcttgccaagttatgtgtcgaattcttccattcac	1447
Db	38436	TGCCAATTTAATCATCATGCAAGGATTTTGCGCAAGTTATGTGTCCAATTTCTTCCATT	38377
QY	1448	acaacactaaacttaattagaactctagaanaatatlltaaatgccaacttcatcgaaaa	1507
Db	38376	ACAACACTAAACTTAATTATGAACCTTAGAATAATTTTAAATGCAACTTTATCGAAAA	38317
QY	1508	aaatttagttatgaanaacaattccagaatlaaacatgagctatataaatttaagataaat	1567
Db	38316	AAATTAGTTATGAACAACATTTCCAGAAATTAACATGAGCTATATTAATTTAAGATTAAT	38257
QY	1568	gaagtaatatltgatatgtatgtaataacatctgattgcggtcaaaaaaacatatctcg	1627

Db 38256 GAAATAATATGATATGATATGTAATAACATATCTGATTGGCTAAACAAAAACATATCTG 38197  
Qy 1628 ataaattgtcatgcagcccatgtcacatgatgtcatcagcttttatttccacat 1687  
Db 38196 ATTAATTGTCATGCAGCCCATGTCACATGATGTCATCAGCTTTTATTTCACAAT 38137  
Qy 1688 aacaaatatataatcacaacaaatagtttgcagattaaatttttgggtgcagct 1747  
Db 38136 AACTAATATATATTCAAAAAATAGTTTGTCAGATTAAATTTTGTGCTGTCAGCT 38077  
Qy 1748 tctccaacctactaaactagtttggaaatgtctctcttcttatttcttcttcttgattc 1807  
Db 38076 TCTCCAACCTACTAACTAGTTTGGAAATGTCCTCTTATTCTTTCTTTCTTGATTT 38017  
Qy 1808 cttatgttttattcatggaatttlaagacgagattgttagtgcgttctcttct 1867  
Db 38016 CTATGTTTTTATTATTAGGAATTTAAGACGGATTGTTAGGTCGTTCTCTCTTTCT 37957  
Qy 1868 tgttttctaaagttaacctttgtaaactcctcccaattagaacagtcaatcatatag 1927  
Db 37956 TGTTTTCTAAAGTTACTTTTGTAAACTCATCTCCCAATTAGACATCATATATAG 37897  
Qy 1928 ttatcctttaaataatgtctagttgataaaaaaaatgaaaaaactgtgtgtatcta 1987  
Db 37896 TTATCTTTTAATATATGCTAGTGTGATAAAAAATGAAAAAATGACTGTGTAGTTCTA 37837  
Qy 1988 ctaatgttctgtgtaaaaaatctgatatattatgaatcctaacttcttgcgtataat 2047  
Db 37836 CTATGTTTGTGTAATAAATCTGATATTATGAATCTAATCAATTTCTTGATCGTATAAT 37777  
Qy 2048 gtgggtttaaatttagtaattttacataaataagaactgtaatgtgtatattgg 2107  
Db 37776 GTGGGTTAAATTAGTAATTTTACATAAATAAGAACTGTAATGTTGATGATATTGGG 37717  
Qy 2108 gaatcagtatataagcttgggtaactactcttggaaataacttgaagatttaactatt 2167  
Db 37716 GAATCAGTATATTAGCTTGGGTAAGTATACTCTGGAATACTGAAATTACTATTT 37657  
Qy 2168 gcaaaatataaatttagtcccgaaaaaatacagacgacgagacacacacataaagcag 2227  
Db 37656 GCAAAATATATATTAGTCCCGAAAAATACAGACGACGGACACGACACATATAAGCAG 37597  
Qy 2228 gtttgaatcttggaaaatttgtlatacacaactatataaataactaagtctgtgg 2287  
Db 37596 GTTGAATCTTGGAATAATTTGTATACATAACCTATATAATACTAATGTTCTGTTGGG 37537  
Qy 2288 ttcaaaagcctttccaagaagctccacttctttaaattcaagqacatttacaatagaata 2347  
Db 37536 TTCAAAAGCCTTTTCAAAAGTTCATTTTTHAATTCAAGACATTTTACATAGGAAATA 37477  
Qy 2348 agttgaagtcataaaaaaataatggtlatacttgaaggttttcttggactaaaagcaca 2407  
Db 37476 AGTTGAGTCATAAAAAATAATGGTTATTGTTGAAGGTTTTTTTGTGATTAAAGCGACA 37417  
Qy 2408 tattaagaagttagtttttctactaccaaatatcaatlaattlaaaacocatgcaacca 2467  
Db 37416 TATTAGAGAGTTAGTTTTTTTTCACIACCAAAATATCAAITTAATTTHAAACCATGCAACCA 37357  
Qy 2468 ttcaaaaaacaatactatlaaagaatataaataacacaaatatlaaatlacacttaaa 2527  
Db 37356 TTCAITAAACAAATACTAATAAAGAAATATAATCACAAAATATTAAATACACTTAAAA 37297  
Qy 2528 ttacataataatttcaaaaacatactaatlaattgaacagaaagaaaaggtlaaatat 2587  
Db 37296 TTACATATAAATTTACAAAAACATCTAATTAATTGAAACAGAAAGCAAAAGTAAAAATAT 37237  
Qy 2588 alcalaanaatqagacatatatacctataaaaaaaaatgagycatatgaagtaataataa 2647  
Db 37236 ATCATAAAAATGAGACATATATCTATAAAAAAAATGAGGCATATGAAGTAATAATAA 37177  
Qy 2648 gagacatgcatgtgaagcattcggttaattaatcgagtcagaagatatatacagtaaatac 2707  
|||||

Db 37176 GAGACATGCATGTAAGCATTCCGTTAAITTAATCGAGTCGAAGATATATATCAGTAATATAC 37117  
Qy 2708 atatgtataatctcggaaaaaagaatatatatatgtgaaataaagaaaagatgaaaatg 2767  
Db 37116 ATATGTATATATTCTGGAAAAAAGAAATATATATTATGAGAAATAAGAAAAAGATGAAATG 37057  
Qy 2768 gaaatgtatataaaaaaagaaggagtggcgacctctgtgtctcg:ccaaaagcaga 2827  
Db 37056 GAAATGATATGATATAAAAAAGAAAGAGTGTCGACTCTGTGCTCGTCCAAAAGCAGA 36997  
Qy 2828 tccaccacagcccccaaaagaagcatgtgagagcctcagcctcaccatctctcatatgat 2887  
Db 36996 TCCAACCAAGCCCCCAAAGAAGCATGATGAGGCTCAGCCTCACCATCTCTCATGGAT 36937  
Qy 2888 tggaaacaaagctaatgatctctcacaacaagaacacgcagcttttctcaatgatctctac 2947  
Db 36936 TCGAACAAGCTAATGATCTCTTCACACAAGAACACGACGCTTTTCTCATGATCTCTCAC 36877  
Qy 2948 catctcatgttagatccacctcccgaaaacccctaattcaacttggacgaaagcagaatgac 3007  
Db 36876 CATCTCATGTTAGATCCACCTCCCGAAACCCCTAATTCTACTTGGACGAAGACGAAGACTAC 36817  
Qy 3008 gatgaagacatgtgattgcgatgaaagagatgcaatgacatgacgtcgcgtcatgcccgtta 3067  
Db 36816 GATGAAGACATGATGCGATGGAAGAGATGCAGTACATGATGCCGTCATGACGCCGTA 36757  
Qy 3068 gacatcgacccctgcacagcgtcccttaagccgaaacgcgcgtaaacgtaagataagcagcat 3127  
Db 36756 GACATCGACCCCTGCCACCGTCCCTTAAGCCGAACGCCGCTAAGGATAAGCGACGAT 36697  
Qy 3128 cctcagaagctgtgtcgtcgtcgagcgtcgaaagatcagcagaaagatccgaatctc 3187  
Db 36696 CCTCAGACGGGTGCTGCTCGTGGGTCGGGAAAGGATCAGGAAAGATCCGAATCTC 36637  
Qy 3188 aagagatcgtgcctgtgtgtgcgaagatgacacagcctccatgctcgacgaagccata 3247  
Db 36636 AAGAGATCGTGCCTGTGCTGTCGGAAGATGACACAGCTTCCATGCTCGACGAAGCATA 36577  
Qy 3248 cgttacaccaagttcttgaacagcgtgagatcttcaagctcactccagatctga 3307  
Db 36576 CCTTACACCAAGTCTTGAAACGGCAGCTGAGGATCTTCAAGCTCTCAGATTCGA 36517  
Qy 3308 gctcctatgtgtcaacccctcttaccttqtlattaccacaactcccaaccctgatlgaact 3367  
Db 36516 GCTCCTATGGCTAACCCCTCTTACCTTTGTTATACCACAACTCCCAACCTGATGAAT 36457  
Qy 3368 acacagaagctcgtctagctagacatlttggtgtcactcctccaaccttttcatgttata 3427  
Db 36456 ACACAGAAGCTCCCTAGCTAGACATTTGGTGTCAITCCTCAACCTTTTTCATGTGAIA 36397  
Qy 3428 tattatataagatgcatataaagatcgcataccaaagatgtatgggttcttaatatatta 3487  
Db 36396 TATTATATATGATGCAATAAGATTCGATCCCAAGATGTATGGGTGTTTAATATTATTA 36337  
Qy 3488 ctctaagataatgatgtacaatgtgtaccaaagttcttatacttgatatacatatcat 3547  
Db 36336 TTCTAAGATATATGATGTACAATTGTGTACCAAGTTTCTTATCTGTATATCATATGCAAT 36277  
Qy 3548 aaataatgtgtgaataaaaaaagagatattgattgtlaaacaaaaaaaagagatattgatt 3607  
Db 36276 AAATAATGTGTAATTAATAAAGAAAGATATTGATTGTAAACAAAAAAGNAGATATTGATT 36217  
Qy 3608 gtttaattagggttltgcatctctgtatgaaagcttggcctgcaaatlaatttccatatat 3667  
Db 36216 GTTAATTAGGTTGATCAATCTGTATGAAAGCTTGGCTGCAAAATTAATTTTCGATAT 36157  
Qy 3668 atatatatatatgagaatatataccaataacttlttaatttgactataaatttqat 3727  
Db 36156 ATATATATATATATGAGAAATATATCAAAATACTTTTAAITTTGACTATAATTTGTAT 36097  
Qy 3728 caattatcgaatctgattgagttgagttatatatgatatagcaaaaagaaaacaccca 3787  
Db 36096 CAATTATCTGAATCTGATGAGTGTAGGTATATATGATGATTAACAAAAAGAAACCAACCA 36037  
|||||







DB 226 ELVPMCTKVD:VTMLKALSYJKFLQVQVKVL 257

RESULT 6

B86402

protein T22C5.19 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: B86402

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.; Schwartz, J.R.; Shinn, P.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Southwick, A.M.; Sun, H.; Tallon,

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: B86402

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-279 <STO>

A:Cross-references: GB:AF005172; NID:g6693022; PIDN:AAF24948.1; GSPDB:GN00141

C:Genetics:

A:Gene: T22C5.19

A:Map position: 1

Query Match 16.6%; Score 175; DB 2; Length 279;  
Best Local Similarity 22.1%; Pred. No. 1.1e-07;  
Matches 60; Conservative 44; Mismatches 63; Indels 104; Gaps 10;

QY 13 SCVSSKSRNSHSPKRSMEPQPHLLMDWKNKANDLLTQEHAAFL----- 57

DB 19 DQSSSK---EERPRDELLGLSLYNGHLHQH-----NNVLSDDHAAFLPDMF 67

QY 58 -----NDPHLLMDPPET-----LIHLDEDEYDEMDAMKEMQYM 94

DB 68 PFGAMPGCGNPAMLDSWDQSHLQ-----ETSLKRLLDVENLCKTNSNCDVTRQELAK 122

QY 95 IAVMQPYDIDPATVPPN-----RRNVRISDDPQIVVA 127

DB 123 SKKKQRVSSSNIVDESNTNMWDQSSISNSDDEKASVTSVKGRTRATKGTATDPQSLYA 182

QY 128 RRRRERISEKIRILKRIVPGAKMDTASMLDEAIRYTKFLKQVRIQ----- 175

DB 183 RKRREKINERLKLQNLVPMNGTKVDISTMLEAVHYVKFLQYIKVCRLKLGFEVFSQNL 242

QY 176 --PHSQI-----GAPMA-NPSYLCYYHN 195

DB 243 IITHVQLSSDDLWMYAPLAYNGLDMGFHHN 273

RESULT 7

T05216

hypothetical protein F1715.70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999

C:Accession: T05216

R:Bevan, M.; Vitale, D.; Liguori, R.; Argiriu, A.; De Simone, V.; Hohnel, J.; Mewes,

submitted to the Protein Sequence Database, July 1998

A:Reference number: 215404

A:Accession: T05216

A:Molecule type: DNA

A:Residues: 1-349 <BEV>

A:Cross-references: EMBL:AL031032

A:Experimental source: cultivar Columbia; BAC clone F1715

C:Genetics:

A:Map position: 4

A:Introns: 177/3; 279/3; 301/3; 323/3

A>Note: F1715.70

Query Match 16.2%; Score 170.5; DB 2; Length 349;  
Best Local Similarity 44.6%; Pred. No. 3.5e-07;  
Matches 33; Conservative 20; Mismatches 18; Indels 3; Gaps 1;

QY 104 DPATVPPN---RRNVRISDDPQIVVARRRRERISEKIRILKRIVPGAKMDTASMLDEA 160

DB 252 DPSKALNLNGKTRASGAATDPQSLYAKRRERINERLRIQNLVPMNGTKVDISTMLEEA 311

QY 161 IRYTKFLKQVRI 174

DB 312 VHYVKFLQIQIKL 325

RESULT 8

G85362

hypothetical protein AT4g30980 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C:Accession: G85362

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Col

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: G85362

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <STO>

A:Cross-references: GB:NC\_001268; NID:g7270000; PIDN:CAB79816.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g30980

A:Map position: 4

Query Match 16.1%; Score 169.5; DB 2; Length 367;  
Best Local Similarity 26.8%; Pred. No. 4.5e-07;  
Matches 61; Conservative 24; Mismatches 68; Indels 75; Gaps 9;

QY 22 SNHSPKRSME-----PQPHLLMDWKNKANDLLTQEHAAFLNDPHHL----- 63

DB 46 SHSPKPKKTEAMNSSLTPSSSPSPH-----LQSPATFDHDDL---HHIFSSIP 94

QY 64 ---MLD--PPET-----LIHLDEDE-----EYDEMDAM----- 88

DB 95 WPSSVLDTPPTSDCAPYTGFGHHADSRNQITMIPLSNHNPNDALEFGFSGSLPFI 154

QY 89 -----KEMQYMAVMPVDIDPATVPPKRRNRNRISDDPQIVVARRRRERISEKI 138

DB 155 POGSGGQTOTQSQATASATTGATAQPTKPKVARRRGQ-ATDPHSIAERLRERIERM 213

QY 139 RILKRIVPGAKMDTASMLDEAIRYTKFLKQVRIQPHSQIGAPMAN 186

DB 214 KSLQELVPNGNKTDKASMLDEIIDIYVKFLQIQVLSMSRLGAASAS 261

RESULT 9

E84634

hypothetical protein At2g24260 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: E84634

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tai

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ve

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: E84634

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-350 <STO>

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 10:27:54 ; Search time 5365.19 Seconds  
(without alignments)  
11856.640 Million cell updates/sec

Title: US-09-548-971B (1)  
Perfect score: 3856  
Sequence: 1 ctctagaccatctactatcc.....caagaggggttcctctagag 3856

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rod:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	3842	99.6	95643	8	F6N15	AF069299 Arabidop
C 2	3842	99.6	197119	8	ATCHRIV1	AL161471 Arabidop.
C 3	3114	80.8	167234	8	AC004708	AC004708 C1C5B11.
C 4	324.6	8.4	367	8	AY027278	AY027278 Arabidops
C 5	172	4.5	68098	8	AB020752	AB020752 Arabidop.
C 6	172	4.5	92491	8	ATF17114	AL353994 Arabidops
C 7	133.2	3.5	90448	8	AC079733	AC079733 Arabidops
C 8	133.2	3.5	90448	24	AC079730	Ac079730 Arabidop
C 9	129.8	3.4	187576	2	AL139811	AL139811 Homo sapi
C 10	128.4	3.3	14867	3	AE001398	AE001398 Plasmodiu
C 11	128.2	3.3	98734	2	PFMAL1P2	AL031745 Plasmodiu
C 12	125.2	3.2	4601	3	DMU11584	U11584 Drosophila
C 13	125.2	3.2	19517	3	DMU37541	U37541 Drosophila
C 14	125.2	3.2	74342	8	AB020742	AB020742 Arabidops
C 15	124.8	3.2	36977	2	AC092304	AC092304 Homo sapi
C 16	124.2	3.2	176174	9	AC007483	AC007483 Homo sapi
C 17	122	3.2	175223	9	AC004617	AC004617 Homo sapi
C 18	121.8	3.2	6598	3	AF205580	AF205580 Plasmodiu
C 19	121.6	3.2	4601	3	DMU11584	U11584 Drosophila
C 20	120.6	3.1	126999	9	AL513328	AL513328 Human DNA
C 21	120.6	3.1	234112	3	PFMALAP2	AL035475 Plasmodiu
C 22	120.6	3.1	67970	3	PFMAL1P3	AL031746 Plasmodiu
C 23	120.4	3.1	104992	2	AC005504	AC005504 Plasmodiu
C 24	120.2	3.1	111122	8	ATF11C1	AL132976 Arabidops
C 25	120.2	3.1	169546	2	AC004157	AC004157 Plasmodiu
C 26	120.2	3.1	160416	2	AL590287	AL590287 Homo sapi
C 27	117.8	3.1	86827	3	PFMAL3P5	AL034556 Plasmodiu
C 28	117.8	3.1	326924	2	AC093082	AC093082 Homo sapi
C 29	117.8	3.1	326924	2	AC093082	AC093082 Homo sapi
C 30	117.8	3.1	326924	2	AC093082	AC093082 Homo sapi
C 31	117.6	3.0	14867	3	AE001398	AE001398 Plasmodiu
C 32	117.6	3.0	86827	3	PFMAL3P5	AL034556 Plasmodiu
C 33	117.2	3.0	175223	9	AC004617	AC004617 Homo sapi
C 34	116.8	3.0	137889	9	AC073269	AC073269 Homo sapi
C 35	116.6	3.0	126999	9	AL513328	AL513328 Human DNA
C 36	116	3.0	143331	9	AC091214	AC091214 Homo sapi
C 37	115.6	3.0	57970	3	PFMAL1P3	AL031746 Plasmodiu
C 38	115	3.0	183584	9	AC012492	AC012492 Homo sapi
C 39	114.8	3.0	136098	9	AC006970	AC006970 Homo sapi
C 40	113.8	3.0	163443	2	AC006280	AC006280 Plasmodiu
C 41	113.8	3.0	318221	2	PFMAL1P3	AL049184 Plasmodiu
C 42	113.4	2.9	15421	3	PFCOMP1RA	X95275 P.falciparu
C 43	112.8	2.9	213530	2	AC090493	AC090493 Mus muscu
C 44	111.6	2.9	149733	2	AC087568	AC087568 Pan trogl
C 45	111.2	2.9	6598	3	AF205580	AF205580 Plasmodiu

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
F6N15/c (1)	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	AF069299.1	GI:3193311	thale cress.	Arabidopsis thaliana	1 (bases 1 to 95643)	Washington University Genome Sequencing Center.	The A. thaliana Genome Sequencing Project	Unpublished (1997)
								2 (bases 1 to 95643)	Ryan, E., Edwards, J. and Page, K.	The sequence of A. thaliana F6N15	Unpublished (1998)



REFERENCE 3 (bases 1 to 95643)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAY-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63108, USA  
e-mail: twilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry: an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 3' cosmid is F5110, 1300 bp overlap. Actual start of this cosmid is at base position 1 of F6N15; actual end is at 1950 F5110.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES  
source

1. .95643  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/clone="F6N15"  
/chromosome="IV"  
/map="unknown"  
complement(829. .4226)  
/gene="F6N15.14"  
complement(join(829. .942,1006. .1086,1156. .1237,1315. .1500,1574. .1675,1776. .1809,1863. .2001,2106. .2230,2325. .3299,3548. .3709,3844. .3987,4073. .4226))  
/gene="F6N15.14"  
/note="contains similarity to breast cancer susceptibility (Brca2)"  
/codon\_start=-1  
/evidence-not\_experimental  
/protein\_id="AAC19315.1"  
/db\_xref="GI:3193333"

gene  
CDS

translation="MSTQWLEPDDSGDGFMEVAGRILOSVDSTPTKALESTAPLPS  
MADLLQGCCKLIAREAMPGEIIMFRTGLGKSVYLKESSIAKAKSILAETYSDLR  
NTNCSIPOMRQVDAETLPLMERTASGKSVPLKESSIAKAMSILGSDKIISDNVLPRE  
SGFGVNSLFOQTASNKKVNVSSAGLARAALIGLEEDLNGFNHYNSSSQHGMS  
GLTHEFDATVVKHSGTPEGQEDYVSGRSEVLNPSLKVPTKFOTAGGKLSVSA  
EALKRARNLGDPELGSFFDDVAGDQFTPEKDERLSDAIINNGSANRGYIAHEEKT  
SNKHTPNFSVSPLEWSSSKQSSVNLLENLASGNIKKFDAVDETDALNATHGLSN  
RSLASDMAVNSKVNFIIPRGQGRPADQPLVDITNRDITAYANKQDSTQKKRLGK  
TVSVSPFKPRPRISSFKTPSKKHALQASSGLSVSCDTLTSKKVLSSTRYPEKSPRYIK  
DFGGMPTATTTRTMSESNQVMQINMYSVSLQTRLELKLFRKWLISLVLPYNMHP  
KRYCDRSFEACMWIWMKLACYDIYPAPKCGNFLTITNVLEELKRYEREVNHGCSA  
IKRILSGDAPASSMAYLCISAINPKTDNSQEAHCSDSCSNVYVELTDGWSMAALD  
VVLTKQLNAGKLFVGOKLRHAYVLISYKLILGAGLSGWATPISPLESFIQPNQAVIS  
STICLLININGTYRAHWADRLGFC"

gene  
CDS

7427. .8228  
/gene="F6N15.13"  
join(7427. .7890,7976. .8041,8120. .8228)  
/gene="F6N15.13"  
/note="coded for by A. thaliana cDNA W43412"

/codon\_start=-1  
/protein\_id="AAC19310.1"  
/db\_xref="GI:3193328"  
/translation="MALALSUSACSPPLRRTTRAGFRTSCSIFANDQRAKKLIE  
SEEDRLRTQKDPKKRDEIVNAIESMTVIGRSSITDDSLSATWRLMTTEKEQLF  
EKAGLFGITAGDVLQVIVNKRILNNVITFPDGVFVRSIDIDIASDQVNFNFNS  
LRCKMELPLPPFGKGFENVYMDGEIRVAKDIRDXYLIVDRAPYMWTESFV"  
8515. .9759  
/gene="F6N15.12"  
join(8515. .8695,8783. .9759)  
/gene="F6N15.12"  
/note="similar to plant chalcone and stilbene synthases"  
/codon\_start=-1  
/evidence-not\_experimental  
/protein\_id="AAC19299.1"  
/db\_xref="GI:3193317"

/translation="MLVSARVEKQKRVAYQKATVLAIGKALPSNVYSGENLVEEY  
EIKDNLSTIKDLQHLCKSTVTKRTYTVMSRETLHKYPELATEGSPITKQREIAN  
VVO MAYEASLVCIKEMGRAVEDITHLVYVSSSERLPGGDLYLSAQIGLSNEQRYV  
YFLGICYGLSLRVAKDIAENNPGRVLLTTSFTVLGFRPNKARPNLVGAALFC  
GAALILIGADPTSESPFELHCAHQOFLPOTGVIGRPLSEEGITFKGRDLPOK  
DNVEEFCKKLVAKAGSCALELNDLFWAVHPGGAIIISGLETKLKLKEKLECSRRAI  
DGNVSSNTIFYIMOKVRDELEKKGTGECEWGLGLAFGPGITFECEFLMRNL"  
11756. .13710  
/gene="F6N15.11"  
join(11756. .11816,11897. .12294,12667. .12732,13158. .1  
13624. .13710)  
/gene="F6N15.11"  
/note="contains similarity to transcriptional activators  
such as Ra-like and myc-like regulatory R proteins"  
/codon\_start=-1  
/evidence-not\_experimental  
/protein\_id="AAC19308.1"  
/db\_xref="GI:3193326"

/translation="MILRQPPPPSPAPQPPQTSPIYVEVAELTWENGOLGHLGPPR  
TASSTKYSTGAGTLESIVDQATRLPNPKPTDELVPWFHRRSSRAAMANDALVPCSN  
VHEQSKPGGVGSTRVSGSCSDGRTPLGGKRRARVAPEWSSGGSORLTMDTYDKRDKI  
ORMKTLOKLVPNSSKTDKASMLDEVIEYLKQLOAYSMSSRMNPPSMPLMAMQOOO  
LOMSLMSNPMGLMGMGMPGLGLDLNSMNRRAASAPNIHANMPPNPLPMNCPSWD  
SSNDSRFQSLPIPDPMSAFLACSTQPTTMEAYSRMALTLYQMOOQLPPSPNPK"  
15299. .20892  
/gene="F6N15.10"  
join(15299. .15951,16806. .17061,17136. .17246,17320. .19  
19156. .19234,19332. .19628,19799. .20001,20363. .20614,  
20746. .20892)  
/gene="F6N15.10"  
/codon\_start=-1  
/evidence-not\_experimental  
/protein\_id="AAC19309.1"  
/db\_xref="GI:3193327"

/translation="MENRHLPSKFFSSASSMAQNOLIDLSHISLYSHSSSSSMAT  
TIPNPSAILRMFSSLSVHORLSHLTVVDPKFVQILLOMLGYIRTKGPCSFIILPDLF  
SSDLPSPLCFKSRGLISRVSESNESERFVFDSTRLFGSGGGERAODSCSVNSLDSY  
VMAEEFLTNVDRFVETMDVDSGAFLRGEESDLGSMVYELMLKAKGYSMERAVANS  
SMGILITTLVDCILRLKLRGLFWYISIDSVKSELLDLNTNHCSPSSSSNQXLGSTRNKQ  
GKTRNNKKPTPEAKSDKNVNLSTKNGKKDQAKLEFNKSRREALTECKKVPRTASTMINDP  
ASAATNKKVYVGLVARKGRTKRRKREKNKSKCTSLBNNGEVANNSVNSAIVKASKCL  
SSCTSANQHPOEYINAQIIEHGSSFECERNRSCTCASNGAANCIFYSGEESRSKAEI  
HVISDLSVDVPAGGSCENVPQKSCCRGDRKEKLIIPNERSRTLDEGSHRIHQF  
REAGYGAASSSEFVSYEMPAVAPMYFSHVSSHLPATDRHLDVDGHNHLYVRQFV  
STVOHRNPSTIEGSHKOVLTRMPMSLDWPMVHSMGLTTAFTCNVDSGILLVDIPEC  
KNKHELGNCEENMMLEDEPVAHTVSQVDYNOYFGGVWYWNPSDHLCTGFSRPPSL  
SDSSMAMWHEAEEMKRSVDWVAFFFFSSYSANGLDSPFAASFCSPHPLGPPNQPLGYV  
PGNEISTKILQAPPTTIEGAGEEVSGTLASLGDVEGNSGDSLPIPLRPITIPNMS  
KSEYKRSYDTKSPNVPTKREHPRIRKPPSPVYLCVPRAPRPPEPSVNSRARRGFP  
TVRSSSSPRHGMGRGWFHDGVNMEEPGAETIYLPWRNKSIAVRPIIQPLGALLQDH  
LIAMSGLDGRDGHVSNVFDIYPDVAFPLQPELLNCPMGESLSLIHENYMSIFEYCI  
YFSWDGVSSNPLNYILLAKVAENMARKPYINMAIKVTRSLQVLMPRSRNTNIFGSSA

[illegible]

RESULT	2		
LOCUS	ATCHR1/c		
DEFINITION	ATCHR1	19719 bp	DNA
ACCESSION	AL161471	Arabidopsis thaliana	DNA chromosome 4, contig fragment No. 1.
VERSION	AL161471.2	GI:7267087	
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		

REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 5139 to 100781; 99482 to 197119)  
Wilson, R., Lamar, B., Stoneking, T., Stumpf, J., Mewes, H.W., Iemcke, K  
and Mayer, K.F.X.  
2 (bases 1 to 6001)  
Zhong, J., Ma, P., Parnell, L.D., Chen, C.N., Chen, E.Y., Mewes, H.W.,

JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 197119)
AUTHORS	EU Arabidopsis sequencing project.
TITLE	Direct Submission
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer

**COMMENT** Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mibs.biochem.mpg.de/proj/thal/> this fragment has an overlap with ATRHRV2 at the 3' end.

FEATURES	Location/Qualifiers
source	1. .197119

```
misc_feature      1537..2290
                  /note="homology to fragments of putative reverse
                  transcriptase"
gene              complement(join(3199..3438,3545..3749,3977..4119,
                  4406..4588,4657..4908))
                  /gene="AT4g00010"
gene              3199..4908
                  /gene="AT4g00010"
exon              complement(3199..3438)
                  /gene="AT4g00010"
                  /number=1
CDS               complement(join(3199..3438,3545..3749,3977..4119,
```

[illegible]

	KIGLY
intron	complement(3439. .3544)
exon	/number=1 complement(3545. .3749) /gene="AT4g00010"
intron	/number=2 complement(3750. .3976) /number=2
exon	complement(3977. .4119) /gene="AT4g00010"
intron	/number=3 complement(4120. .4405) /number=3
exon	complement(4406. .4588) /gene="AT4g00010"
intron	/number=4 complement(4589. .4656) /number=4
exon	complement(4657. .4908) /gene="AT4g00010"
misc_feature	/number=5 5139. .6001

gene	complement(5967..6080) /gene="AT4g00020" /number=1 5967..9364 /gene="AT4g00020"
exon	complement(join(5967..6080,6144..6224,6294..6375, 6453..6638,6712..6813,6914..6947,7001..7139,7244. 7463..8437,8686..8847,8982..9125,9211..9364))
CDS	

```
Contains Prokaryotic membrane lipoprotein lipid attachm
site AA475-485"
/codon_start=1
/product="putative BRCA2 homolog"
```

```

/translation="MSTWOLFPPDSSGDFRMEVAGRILQSVSDSTFPTKALESTIA
MADLLGCGCKLILAREEAMPEGIPEMFTGSLGVSVLKESSIAKASILAEKVITY
NTNCSIPQMRQVDTAETLPMERTASGVSPLEKSSIAKAMSLGSKITIDSDNV
SGFVYNSLIPOTASNKKVAVSSAGLARAKALLGLEODLNGFNHYNOSSSSSOQ
GLKTHEEFDATVVKHSGTPGQYEDVYSGKRESELPISKVPTKFTQTAGSKSL
EALKRARNLLGDPELGSEFFDDVAGGDQFFTPENDERLSDIAINRGSANRGYIAH
SNKHTPNSEFVSPLMSSSKQFSSVNLLENLAGGN:IKKFDAAVDETDICALNATHG:
RSIADSMAYKNSXVNGEPIPRGROPGRADOPYDITNRDITAYAKKODSTOKK:
TVSVPFPRKPRISSFKTPSKHALQASSGSLVSCDTLTSKKVLSIRYPEKSPR
IDFGMHPATATRTMSFESNOVMOIMYSVMSLQTRLELKFPCMSLVLPPYNNA
KKVCDRSFEACMMIVMKLACYDIYYPAKRCGNPLTITNVLEELKRYEREVNHG:
IKRILSGCAPASSMPLYCISAINPKTIDNDSQEAHCSDCSSNVKVELTDGYSMN
VVLTKOLNAGKLFVGGQKLEHAYVLSYLKILGAGLSGWAITPSPLESFIQPNFQ:
STTCLLINNGTYRAHWADRLGFCR"

```

intron	complement(6081. .6143) /number=1
exon	complement(6144. .6224) /gene="AT4g00020" /number=2
intron	complement(6225. .6293) /number=2
exon	complement(6294. .6375) /gene="AT4g00020" /number=3
intron	complement(6376. .6452) /number=3
exon	complement(6453. .6638)

/gene="AT4g00020"  
/number=4  
complement(6639. .6711)  
/number=4  
complement(6712. .6813)  
/gene="AT4g00020"  
/number=5  
complement(6814. .6913)  
/number=5  
complement(6914. .6947)  
/gene="AT4g00020"  
/number=6  
complement(6948. .7000)  
/number=6  
complement(7001. .7139)  
/gene="AT4g00020"  
/number=7  
complement(7140. .7243)  
/number=7  
complement(7244. .7368)  
/gene="AT4g00020"  
/number=8  
complement(7369. .7462)  
/number=8  
complement(7463. .8437)  
/gene="AT4g00020"  
/number=9  
complement(8438. .8685)  
complement(8686. .8847)  
/gene="AT4g00020"  
/number=10  
complement(8848. .8981)  
complement(8982. .9125)  
/gene="AT4g00020"  
/number=11  
complement(9126. .9210)  
complement(9211. .9364)  
/gene="AT4g00020"  
/number=12  
12565. .13366  
/gene="AT4g00030"  
12565. .13028  
/gene="AT4g00030"  
/number=1  
join(12565. .13028,13114. .13179,13258. .13366)  
/gene="AT4g00030"  
/note="coded for by A. thaliana CDNA W43412  
contains EST gb:W43412"  
/codon\_start=1  
/product="predicted protein of unknown function"  
/db\_xref="GI:7267090"  
/translation="MALALSLSACSPPLRTRRAGFRTSCSIFANPAQRAKRLLE.I  
SEDRGLRTQKDPKKRDEIVNAIESMTVIGRSSITDDLSATWRLMTTEKEQLFI  
EKAGLFGTLAGDVLOVIDVKNRLINNVITPPDGVEFVRSDDIDIASPQRYNFRNSAV  
LRGKNWELPLPPFGKGFENVMDSIRVAKDIRGDIYLVDRAPYNMTESFV"  
13029. .13113  
/gene="AT4g00030"  
/number=1  
13114. .13179  
/gene="AT4g00030"  
/number=2  
13180. .13257

Query Match 99.6%; Score 3842; DB 8; Length 197119;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ccattactatccggtgttgacccttaagcctttgaagactagataatgcaaat 67  
|||||

Db 44954 CCATCTACTATCCGGTTGTTGACCCCTTAAGCCTTTTGAGAGCTACTAGAAATATGCCAAT 44895  
QY 68 accatagtcacatataccatcccttcttctgttgtaactgaacattctaatttglaaa 127  
|||||  
Db 44894 ACCATATGTCATATCCATCCCTTTCTTTGTTGTAAGTGAACATTCTAATTTGTAAAA 44835  
QY 128 gaaaaaacctatgttaataataccqtaggcaaaaaatatctcatcataatattt 187  
|||||  
Db 44834 GAAAAAACCTTATGTTAATATACCCGTAGGCCAAAAAATATCTCATCATATAAATTT 44775  
QY 188 tatataagattatacatctctcgttgtaagagttactccaattgcaagtggtatta 247  
|||||  
Db 44774 TATATAGATATATACATTCTCTGTTGTAAGAGTACTCCAATTGCAAGTGTGTATTA 44715  
QY 248 actaataaaagagcaagagtagaagcctataataattgatgttgcatagtcgtla 307  
|||||  
Db 44714 ACTAATAAAAAGAGCAAGAGTAGAAGCTTATATTAATGTGATGTGCATAGTACTGTA 44655  
QY 308 tatgttgatgaataatacaagatgaacatlaatgaatgaacagggtatttgcctg 367  
|||||  
Db 44654 TATTGTTGATGAATATACAAGATGAACATTAAATGCAATGAACGGGATTTTGTCITG 44595  
QY 368 aactcatlaaaggcaatgtgaaagaagatgtgagctcatttgaaaattatctct 427  
|||||  
Db 44594 AACTCATTAAGGCAATGTGAAGAAAGATGTGAGGTCCTCATTTGAAAAATTATCTTCT 44535  
QY 428 agcttgcgatlltaaatctatgaatgaacgcaacatataagaatttcatgtgacaa 487  
|||||  
Db 44534 AGCTTGTGATTTAATCTATGAATGAACCGCAACATATAGAATTTGATGIGACAA 44475  
QY 488 cgacattagacggtatccttaattagaccgattatgaataataataataataata 547  
|||||  
Db 44474 CGACATTAGACGGTATCTTAATTAGACCGATTAAATTAATTAATTAATTAAT 44415  
QY 548 tagtgatgattataagttactatccactgaagatttaaaacaatggcaataccttaa 607  
|||||  
Db 44414 TAGTGATGATTATAAGTTACTATCCACTTGACAAATTAACAAATGCGCAATACCTTAA 44355  
QY 608 tgcgaagaagccgtcccccacttcgttaatgagttatgaggagagatcctgttaac 667  
|||||  
Db 44354 TGTCGAAGAAGCCGTCGCCACTTCGTGTAATGAGATTATGGGGAGAGATCTCTTAAT 44295  
QY 668 cgtcaataaaacaacttaagaactgaatltgacacccaaaaatcataagaagaacgttg 727  
|||||  
Db 44294 CGTCAAAATAAAACAACCTTAAGAACTAGAATTGACACCAAAAATCATTAAGAGAACGTG 44235  
QY 728 aagaagtcattatcgtatccagctcatatttccagctaagaatcaaatcaaggccgttg 787  
|||||  
Db 44234 AAGAAGTCATTATTCGTATCCAGCTCATATTTCCTAGCTAAGATCAAAATCAAGCCGTTG 44175  
QY 788 aaaggcttqtaagaataatgtcgaagaacacgctggggttagaagaagaagaataag 847  
|||||  
Db 44174 AAAGGGCTGTGAAGAAATGTGCAAGAAACCGTGGGTTTAGAAGAAAGACAGAAATAG 44115  
QY 848 aagaacaatgatgttaaatltgcctatttqgtgtatagagtgtaaaaagaagaagag 907  
|||||  
Db 44114 AAGAACAATGATGTTAATTCCTATTGCTATTTGGTGTATAGAGTGTGCAAAAAGAGAGAG 44055  
QY 908 agaagaataatagtcacaataatgagcactaaaaatgyagacatgtgtgagtaactat 967  
|||||  
Db 44054 AGAAGAATAATTAGTCAAAATATAGAGCACTAAAAAATGCAAGACATGCTGTGAGTAAC 43995  
QY 968 tacaagagcacttatgtcttccctatgcaatgatataccaaacgaagtgaacgctcct 1027  
|||||  
Db 43994 TACAAGAGCAGACTTATGCTTCCTTATGGCAATGATATCCAAACCAAGTGCACGCTCCT 43935  
QY 1028 ttcttgccctaatttcgttaagatcctcctccttctcgttccttaggaaaaaaccttagaa 1087  
|||||  
Db 43934 TTTTGCCCTAATTTGTAAGGTCCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 43875  
QY 1088 tttaatcccttqcttctgactcctgcttcttgaagtaacaaatgatttgaaccacaactaatt 1147  
|||||  
Db 43874 TTAAATCCCTGTCTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 43815

QY 1148 tcttcataccttctggtcctatagagatttgccttatatgcttctctgtattgctccgc 1207  
Db 43814 tctctatctcttgggtctatagagatttgccttatatgcttctctgtattgctccgc 43755  
QY 1208 acgtacgtatacgaatttcaatggttatatacaaggtttatataaactagcacaaatgagc 1267  
Db 43754 acgtacgtatagcaatttcaatggttatatacaaggtttatataaactagcacaaatgagc 43695  
QY 1268 ccatagaatttctgtagcgaataaaggtagaataatctgagctcttcaaacgcgaatatata 1327  
Db 43694 ccatagaatttctgtagcgaataaaggtagaataatctgagctcttcaaacgcgaatatata 43635  
QY 1328 taatttgcctgcaaaactagccttcaatcagatcctaatgatatatttcttcaatttccct 1387  
Db 43634 taatttgcctgcaaaactagccttcaatcagatcctaatgatatatttcttcaatttccct 43575  
QY 1388 tgcgaataatcaatcacatgcacgagatttctgcaagcttatgctgcgaattcttccattcac 1447  
Db 43574 tgcgaataatcaatcacatgcacgagatttctgcaagcttatgctgcgaattcttccattcac 43515  
QY 1448 acaacactaaacttaataagaacctaggaatacttcaaaatgacaacttatacgaanaa 1507  
Db 43514 acaacactaaacttaataagaacctaggaatacttcaaaatgacaacttatacgaanaa 43455  
QY 1508 aaatttagtctatgaaacaaatccagaatccaacatgagctatataatttaagataaaat 1567  
Db 43454 aaatttagtctatgaaacaaatccagaatccaacatgagctatataatttaagataaaat 43395  
QY 1568 gaagtaatatltgatatgtatgtaataacatactgattgcggtlaaaaaaaacatatctg 1627  
Db 43394 gaagtaatatltgatatgtatgtaataacatactgattgcggtlaaaaaaaacatatctg 43335  
QY 1628 attaaattgtctatgcagggcccatgtcactatgattgtcatcagcttcttatttccacaat 1687  
Db 43334 attaaattgtctatgcagggcccatgtcactatgattgtcatcagcttcttatttccacaat 43275  
QY 1688 aactaatatatactcaaaaaaagattctgtcagaatcaaatcttcttctgtgtcagctc 1747  
Db 43274 aactaatatatactcaaaaaaagattctgtcagaatcaaatcttcttctgtgtcagctc 43215  
QY 1748 tctccaacctactaaactagtttgaatgtctctctcttatttcttcttcttctgatt 1807  
Db 43214 tctccaacctactaaactagtttgaatgtctctctcttatttcttcttcttctgatt 43155  
QY 1808 cctatgttttcttattatggaatttcaagaagattgttcaagtcgtctctctctctctct 1867  
Db 43154 cctatgttttcttattatggaatttcaagaagattgttcaagtcgtctctctctctctct 43095  
QY 1868 tgtttcttaaggtacttctgttaaacctcatctctcccaattagacagctcaacatatag 1927  
Db 43094 tgtttcttaaggtacttctgttaaacctcatctctcccaattagacagctcaacatatag 43035  
QY 1928 ttatcttttaataatactcagttctgataaaaaaaatgaaaaaatacctgtgtgattctca 1987  
Db 43034 ttatcttttaataatactcagttctgataaaaaaaatgaaaaaatacctgtgtgattctca 42975  
QY 1988 ctaatgttctgttaaaaaatcgtatataatgaactcaatcttcttctgactgataat 2047  
Db 42974 ctaatgttctgttaaaaaatcgtatataatgaactcaatcttcttctgactgataat 42915  
QY 2048 gtgggttaaatcttagtaatttctacataaagaagctgtaagtctgatatattggg 2107  
Db 42914 gtgggttaaatcttagtaatttctacataaagaagctgtaagtctgatatattggg 42855  
QY 2108 gaatcaglatatagctgtgggtaactatactctggaataactggaagattcaactatt 2167  
Db 42854 gaatcaglatatagctgtgggtaactatactctggaataactggaagattcaactatt 42795  
QY 2168 gcaaaatataatctagctccgaaaaaatacagaagcagggacagacaacatatagaag 2227  
Db 42794 gcaaaatataatctagctccgaaaaaatacagaagcagggacagacaacatatagaag 42735

QY 2228 gttgaatcttggaaaatttctgtatacataaacctataaataactaatgtctgttggg 2287  
Db 42734 gttgaatcttggaaaatttctgtatacataaacctataaataactaatgtctgttggg 42675  
QY 2288 ttcaaaagcctttcaaaagttcccatcttcaaaatcgaagacatttcaataggaata 2347  
Db 42674 ttcaaaagcctttcaaaagttcccatcttcaaaatcgaagacatttcaataggaata 42615  
QY 2348 agttagcataaaaaaataatgattatttctglaaagtttcttctgtatlaaacgcaca 2407  
Db 42614 agttagcataaaaaaataatgattatttctglaaagtttcttctgtatlaaacgcaca 42555  
QY 2408 tatlaagaagttagtttcttccactaaccaatatcaatlaattlaaaccaatgcaacca 2467  
Db 42554 tatlaagaagttagtttcttccactaaccaatatcaatlaattlaaaccaatgcaacca 42495  
QY 2468 ttcatnaaacatlaactatlaagaagatataaataatcaaaaatattlaaaccttaaaa 2527  
Db 42494 ttcatnaaacatlaactatlaagaagatataaataatcaaaaatattlaaaccttaaaa 42435  
QY 2528 ttcatataaaatttacaanaacatlaattlaattgaaacgaaagaaagttaaaatat 2587  
Db 42434 ttcatataaaatttacaanaacatlaattlaattgaaacgaaagaaagttaaaatat 42375  
QY 2588 atcataaaatgagacatatatcctataaaaaaaatgagcagatatacgaatataat 2647  
Db 42374 atcataaaatgagacatatatcctataaaaaaaatgagcagatatacgaatataat 42315  
QY 2648 gagacatgcatgtaagcattcgtgttaattaatcagagctcaagaatatatacgaataac 2707  
Db 42314 gagacatgcatgtaagcattcgtgttaattaatcagagctcaagaatatatacgaataac 42255  
QY 2708 atatgttatatttctgaaaaagaatatataatatgagaataaagaatgaaaatg 2767  
Db 42254 atatgttatatttctgaaaaagaatatataatatgagaataaagaatgaaaatg 42195  
QY 2768 gaaaatggtatgtataaaaaaagaagaggtgtgagactcttctgtctgtccaaaagcaga 2827  
Db 42194 gaaaatggtatgtatataaaaaaagaagaggtgtgagactcttctgtctgtccaaaagcaga 42135  
QY 2828 tccaaccacagcccccacaaagaagcatgtgagacctcagacctcacatctctcatgagat 2887  
Db 42134 tccaaccacagcccccacaaagaagcatgtgagacctcagacctcacatctctcatgagat 42075  
QY 2888 tggaaacaagcctaagtatctctctcacacaagaacacagcagcttctctcaatgattctcac 2947  
Db 42074 tggaaacaagcctaagtatctctctcacacaagaacacagcagcttctctcaatgattctcac 42015  
QY 2948 catctcatgtttagatccacctcccgaaaaccctaattcaacttggacggaagcagaagat 3007  
Db 42014 catctcatgtttagatccacctcccgaaaaccctaattcaacttggacggaagcagaagat 41955  
QY 3008 gatgaagacatgagatgcgagtgaagagagatgcaqtacatgatacgcgtcatcagcccgta 3067  
Db 41954 gatgaagacatgagatgcgagtgaagagagatgcaqtacatgatacgcgtcatcagcccgta 41895  
QY 3068 gacatcgacccctgccaagcaggtccctcaagcgcgaacccgcgttaacgtaaqataagcagat 3127  
Db 41894 gacatcgacccctgccaagcaggtccctcaagcgcgaacccgcgttaacgtaaqataagcagat 41835  
QY 3128 cctcagaagcgtgtgtgtctcgtctgagcgtcgggaagagatcagcgaaagatccgaattctc 3187  
Db 41834 cctcagaagcgtgtgtgtctcgtctgagcgtcgggaagagatcagcgaaagatccgaattctc 41775  
QY 3188 aagagatcgtgacctgtgtgtgtggaagatgagcacacagcttccatgctcgaagaaagcata 3247  
Db 41774 aagagatcgtgacctgtgtgtgtggaagatgagcacacagcttccatgctcgaagaaagcata 41715  
QY 3248 cgttacaccaagctctctgaacgcaggtgagagattctcagacctcaactcagaattgga 3307  
Db 41714 cgttacaccaagctctctgaacgcaggtgagagattctcagacctcaactcagaattgga 41655  
QY 3308 gctccatggtcaacccctctacacttctatraccaaaactcccaacctgagtaact 3367



|||||  
Db 41654 GCTCCTATGGCTAACCCCTCTTACCTTTGTATTACACAACTCCCAACCCCTGATGAAGT 41595  
QY 3368 acacagaagctcgctagctagacattggtgtgtcatcctcccaaccttttcattgtgata 3427  
Db 41594 ACACAGAAGCTCGCTAGCTAGACATTTGGTGTCACTCCTCAACCTTTTTCATGTGATA 41535  
QY 3428 tattatataagacataaagaatcgcaccgaagtgtatggtgtgttttaattattta 3487  
Db 41534 TATTATATATAGATGCATAAAGATTCCGATCCAAGATTGTATGGGTGTTTAAATATTATTA 41475  
QY 3488 ttctaagatatatgatagtacaattgtgtaccaagtttcttatacttgatatcatatgcat 3547  
Db 41474 TTCTAAGATATATGATGTACAAATTGTGTACCAAGTTTCTTTATCTTGATATCATATGCAT 41415  
QY 3548 aaataaattggtgaataaaaaagaagatatgtatgttaacaaaaaaagaagatatgtatc 3607  
Db 41414 AATTAATTGGTGATATAAAGAAGATATTGATTTGTAACAAAAAAGAAGATATTGATT 41355  
QY 3608 gtttaattaggtgtgtatcatctctgtatgaaagctttgacctgcaaatlaatttcgatat 3667  
Db 41354 GTTAAATTAGGGTTTGATCATCTCTGTATGAAGCTTTGGCCTGCAATTATTTTCATAT 41295  
QY 3668 atatatatatatgagaatatatatatccaataactttttaaatttgactataattgtat 3727  
Db 41294 ATATATATATATATGAGAAATATATATCAATACTTTTAAATTGACTATATATTGTAT 41235  
QY 3728 caattatctgaatctgtagagtgtagttatatatgtatgagcaaaaaaagaacaacca 3787  
Db 41234 CAATTATCTGAATCTGTAGTGTATATATGATTAGCAAAAAAGAAACCAACCA 41175  
QY 3788 ttataacgcacctacatlaaaaaatcatcccaagaagaaccatcctcaagaggttc 3847  
Db 41174 TTATTACGCACCTACATTAATAATCATCCCAAGAAAGAACCATCTCTCAAGAGGGTTC 41115  
QY 3848 cc 3849  
Db 41114 CC 41113  
RESULT 3  
AC004708/c AC004708 167234 bp DNA PLN 21-MAY-1998  
DEFINITION C1C5B11.1 check: 4870 from: 1 to: 167234, complete sequence.  
AC004708  
VERSION AC004708.1 GI:3150006  
KEYWORDS  
SOURCE htg.  
ORGANISM thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 167234)  
AUTHORS Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.  
TITLE A. thaliana YAC C1C5B11, NOR of chromosome II  
JOURNAL Unpublished  
COMMENT 2 (bases 1 to 167234)  
REFERENCE 2 (bases 1 to 167234)  
AUTHORS Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.  
TITLE Direct Submission  
JOURNAL Submitted (21-MAY-1998) Applied Biosystems Division of  
Perkin-Elmer, 850 Lincoln Centre Dr., Foster City, CA 94004, USA  
YAC C1C5B11 was sequenced by ACGT, a member of the Cold Spring  
Harbor Arabidopsis Genome Sequencing Consortium. C1C5B11 was  
supplied to us by Dr. Craig S. Pikaard and stated to encode the  
rdna repeats of chromosome IV. C1C5B11 does not align perfectly  
with BAC F5110 which extends into the sub-rDNA region covered by  
C1C5B11 and which has been shown to carry chromosome IV markers  
m151 and m1204 (U. Johanson, pers. comm.). The sequence of C1C5B11  
is not complete. There is a gap between a 12.1 kbp contig and this  
submitted sequence. This is in the region of the rdna repeats and  
the sequence presented here contains several of these rdna repeats.  
For further information, please see http://www.cshl.org/arabweb.  
FEATURES  
Location/Qualifiers

source i. 167234  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="2, see COMMENT"  
BASE COUNT 50750 a 31895 c 33754 g 50835 t  
ORIGIN  
Query Match 80.8%; Score 3114; DB 8; Length 167234;  
Best Local Similarity 88.2%; Pred. No. 0;  
Matches 3387; Conservative 0; Mismatches 455; Indels 0; Gaps  
QY 8 ccactactatccggtgttgacccttaaaagcttttgaagactactagaataatgcaat 67  
Db 70364 CCACTACTATCCGGTGTGTAATATTATTCTTGTGATCAAAATACATATCCATCTACTA 70305  
QY 68 accatatgcccataccatcccttcttcttcttcttcttcttcttcttcttcttcttcttctt 127  
Db 70304 TCCGCTGTGTGATATCCATCCTTTCTTTGTTGAACTGAAATCTAATTGTGAAAA 70245  
QY 128 gaaaaaaccttatgttaatatatcacgctagcgdaaaaaaataatctcatatataatctt 187  
Db 70244 GAAAAAACCTTATGTAATATATCACCGTAGCAAAAAAATATCTCATATATAATTT 70185  
QY 188 tattataagattatacatctctctcgtgtgaagagttactccaattlgcaagtgtgtatta 247  
Db 70184 TATTATAAGATTATACATCTCTCGTTGAAGAGTTACTCCAAITGCAAGTGTGTATTA 70125  
QY 248 actaataaaaaagagcaaaagtaggaagcttataatlaattgtatgttgcatagtactgta 307  
Db 70124 ACTAATAAAAAAGGACGAAGTAGAAGCTTATTAATTAATTGATGTTGCATAGTACTGTA 70065  
QY 308 tattgttgatgaataatacaagtagaacaataatgcatgaaacgsgtatcttgccttg 367  
Db 70064 TATTGTGATGAATATAACAAGAAATTATTAAGTGTGCATAGTACTGCTATATGTTGAT 70005  
QY 368 aactcaattaaagccaatgtgaaaagaagtgtgaagttcattttgaaatttaacctct 427  
Db 70004 GAATATAACAAGCCAAATGTGAAGAAGAGATGTGAGGCTTCATTTTGAATAATTATCTCT 69945  
QY 428 agccttctgcatttlaaatctatgaaatgaaacgcaacatatagaatlttcattgtgacaa 487  
Db 69944 AGCTTTGTGCATTTTAAATCTATGAATGAACGCAATATAGAAATTTCAITGTGACAA 69885  
QY 488 cgaacatttagacggtatcttlaattagaccgaattaatagtaataacttataatataat 547  
Db 69884 CGACATTTAGACGGGTATCTTAATTAGACCGGTTAATTAGTAATATACCTTATATATAAT 69825  
QY 548 tagtggatattataaagtttacttatccacttgaaatlttaaaacaatgagcaataacctaa 607  
Db 69824 TAGTGGTGAATTATTAAGTTACTTATCCACTTGACAAATTTAAACAAATGGCAATACCTTAA 69765  
QY 608 tctcgaagaagaagccgcccactctgtgttaatgagttatgaggaagagatccctgttaaat 667  
Db 69764 TGTCGAAGAAGAGCCGTCGCCACAGAAATTTAAACAATGGGCAATACCTTAATGTGAAAGA 69705  
QY 668 cgtcaataataacaacttaagaactagaatltgacaccaaaaatcatlaaagagaagcttg 727  
Db 69704 AGCCGTCGCCACCAACTTAAGAACTAGAAATGTGAACCAAAAAATCATATAAGACAACGTTG 69645  
QY 728 aagaaglcatttatcgatccagctcatatltccttagctaaagatcaaatcaagcgcttg 787  
Db 69644 AAGAAGTCATTTATTCATCCAGCTCATATTTTCCTAGCTAAGATCAAAATCAAGGCCGCTTG 69585  
QY 788 aaagagcttctgaagaatgtcgaagaacccgltgaggttttagaagaagaagacaagaatag 847  
Db 69584 AAAGGCTTGTAGAAGAAATGTCGAAGAAACCGTGGGGTTAGAGAAAGACAGAATAATAG 69525  
QY 848 aagaacaatgatgttaaatgtcctatttgggttataggaagctgtcgaagaagagagag 907  
Db 69524 AAGAACATGATGTAAATTGCCCTATTTTGTGTATAGAGTTGTCAAAAAGAGAGAGAGAG 69465



[illegible]

Db	68384	CTAATGTTTGTGTAAAAAATCTGATATTATGAATCTAATCAATTTCTTTGATCGTATAAT	68325
QY	2048	gtggttlaaatttagtaatttttacataaataagaactgtaagtgtacataattggg	2107
Db	68324	GTGGCTTAATTTAGTAATTTTTCACATAAATAAGAACTGTAATGTGATGATATTGGG	68265
QY	2108	gaatcagtatattagcttgggtaactatactctctgaaatacttgaagatlttaacttlt	2167
Db	68264	GAATCAGTATATTAGCTTGGGTAGAACTGTAATGTGATGATATTGGGAATCAGTAT	68205
QY	2168	gcacaaattataatttagtcccgaaaaatacagacgacggaacagacacacatalaagcag	2227
Db	68204	ATTAGCTTGGGTTTGTAGTCCCGAAAAATACAGACGACGGACACGACACATATAAGCAG	68145
QY	2228	atttgaatcttggaaaaatttctgtatataaacctalatataataactaatgttctgtctggg	2287
Db	68144	GTTGAATCTTGGAAAAATTGTATACATAACCTATATAAATACTAATGTTCTGGTTGGG	68085
QY	2288	tccaagaagcctttccaaaagttcccatcttttaaatccaagacactttacataagaaala	2347
Db	68084	TTCAAAAGCCTTTTCCAAAGTTCOCATTTTAAATTCGAAGCATTTTACATAGCAATA	68025
QY	2348	agttgaagtcataaaaaataatgtgtatlttqlaaggttttlttggatlaaaacgcaca	2407
Db	68024	AGTTGAGTCATAAAAAATAATGTTATTTTGTAAGGTTT1TTTTCATTAAACGCACA	67955
QY	2408	taltgaagaagltagtttlttccactaccaaalatacaattaatlttaaaaccatgcaacca	2467
Db	67964	TATTAGAAGTTAGTTTTTTAAAGGTTT1TTTTCATTAAACGCACATATTAGAGG	67905
QY	2468	tccaataaacactactataaagaatataaataatcacacaaatataaataacacttaaa	2527
Db	67904	TTAGTTTTTTTTACTATTAAAGAAATATAATTAATCACAAAAATATAATACACTTAAAA	67845
QY	2528	ttacataataaatctacaacaacactaatlaattgaaacugaaggaauagqlaaatat	2587
Db	67844	TTTACATATAAATTTCACAAAACATCTAATTAAATGGAACAGAAAGGATAAATATAT	67785
QY	2588	atcataaaatgagacatatatctcctataaaaaaaaacgaagcatatgaaqtaataataa	2647
Db	67784	ATCATAAATGAGACATATATCCTATAAAAAAAAATGAGGCATATGAAGTAAATATA	67725
QY	2648	gagacatgcatgtlaagcaltcgttlaattatcagtcagtcnaagatatataccagtaaatac	2707
Db	67724	GAGACATGCATGTAAGCATTCGGTTAATTAATCGAGTCAAAAGATATATACAGTAAATAC	67665
QY	2708	atatgtgtatatlttcttggaanaaagaatatatatattgaaanaaagaagaatqaaatg	2767
Db	67664	ATATGTGTATATTTCTGCAAAAACGAGTCAAACATATATATCAGTAATAATACATATGTAT	67605
QY	2768	gaaaaatgatatgtalataaaaagaaggaagtgctgctgactcttgytgcctcgtuccaaaagcaga	2827
Db	67604	ATTCTCGAAAATATATAAAGAAGAGAGTGTGCGACTCTTGCTGCTCCAAAAGCACA	67545
QY	2828	tccaaccacacagcccccacaaaagaagcatgtatgagagctcagcctcaaccatctcctcatgat	2887
Db	67544	TCCAACCACAGCCCCCAAAAGAAGCATGATGAGCCTCAGCCTCACCACATCTCCTCATGAT	67485
QY	2888	tgggaacaaagctaatagtatcttctcacacaagaacacgcagccttctcaaatgatactcac	2947
Db	67484	TGGGAACAAAGCTAATGATCTTCTCACACAGAAACACGACGCTTTTCTCAATCATCTCCAC	67425
QY	2948	catctcatgttagatccacactcccgaaaacccctaatltcacttggacgaagaacgaagaagtac	3007
Db	67424	CATCTCATGTGTAGATCCACCTCCCGAAACCCCTAATTCACCTTGACGGAAGACGAAGAGTAC	67365
QY	3008	gatgaagacacatgcatgcatgaagaagagatgcagfacatgatgcgcgcatgcagcccgta	3067
Db	67364	GATGAAGACATGATGATGCGATGAATAATTCACCTTGACGGAAGACGAAGAGATGAAGACA	67309
QY	3068	gacatgcacccctgccacagctcccttaagccgaaccgcgtaacglaagataaagcagcat	3127

[illegible]

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 367)
AUTHORS	May, B. P., Simorowski, J., Arroyo, J. M., Shen, R., Healy, J., Reiner, A. McCombie, W. R. and Marienssen, R. A.
TITLE	Direct Submission
JOURNAL	Submitted (07-FEB-2001) Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
FEATURES	Location/Qualifiers
Source	1. 367
	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"
	/note="transgenic line GT6209"
	ecotype: Landsberg"
BASE COUNT	105 a 73 c 50 g 134 t 5 others
ORIGIN	
Query Match	8.48; Score 324.6; DB 8; Length 367;
Best Local Similarity	94.98; Pred. No. 7e-38;
Matches 333; Conservative	0; Mismatches 18; Indels 0; Gaps
QY	385 gtgaaagaagatgtgaggtctcatttggaaaattatcctctagcttgcgatttaa 444
Db	362 GAGAGAGAGAGATGTGAGGCTCTCATTTGAAAATTATCTCTAGCTTTGCGATTTCG 303
QY	445 atctatgaatgaacgcgaacacatatagaatttcattgtgacaaacgacatttagacggtat 504
Db	302 TTGAGNGAGATGAAACGCAACATATAGAAATTTCATGTGGACACACGACATTAGACGGTAT 243
QY	505 cttaattagaccgatttaattagtaataatactatatalatalaataatagtggtattataagt 564
Db	242 CTTAATTAGACCGATTATTATTAGTAATATATTCTATATATATTATTAGTGGTATTATAAGT 183
QY	565 ttactatccacttgaagaatttaacaacaggggcaatcccttaa-cgtcgaagaagagccgtc 624
Db	182 TTACTTATCCACTTGAGAAITTTAAACATATGCCCAATACCTTAATGTCCAAAGAAGCCGTC 123
QY	625 cccactcgtgtcaatgagttatgtggggagagatccctgttaaatcgtccaataaacaact 684
Db	122 CCCACTCGTGTAATGAGTTATGGGGAGAGACATCCTGTTAATCGTCAATAAACAACACT 63
QY	685 taagaactagaaaattgcacacccaaaataatcataaagaagaacgcttgaagaagtc 735
Db	62 TAAGAACTGAATTTGACACCAAAATCATTAAGAGAACGACGGTANAGC 12
RESULT	5
AB020752/c	AB020752 68098 bp DNA PLN 25-OCT-2000
LOCUS	Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MTH16.
DEFINITION	AB020752
ACCESSION	AB020752.1 GI:3985955
VERSION	
KEYWORDS	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui pl clone:MTH16.
SOURCE	
ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (sites)
AUTHORS	Kaneko, T., Kato, T., Sato, S., Nakamura, Y., Asamizu, E., Kotani, H., Miyajima, N. and Tabata, S.
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. IX. Sequence features of the regions of 1,011,550 bp covered by seventeen P1 and YAC clones
JOURNAL	DNA Res. 6 (3), 183-195 (1999)
MEDLINE	99397451
REFERENCE	2 (bases 1 to 68098)
AUTHORS	Nakamura, Y.
TITLE	Direct Submission
JOURNAL	Submitted (03-DEC-1998) to the DDBJ/EMBL/GenBank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan

COMMENT

(E-mail: ynakamu@kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)  
Address for correspondence: kaos@kazusa.or.jp  
For the latest information on annotation of this clone, please see [http://www.kazusa.or.jp/kaos/cgi-bin/agd\\_graph.cgi?c-mth16](http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c-mth16)  
Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.  
The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://combio.ornl.gov/Grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi>).  
Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).  
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 3' clone is MYH9.

FEATURES

Source

CDS

CDS

CDS

```
1. 68098
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MTH16"
/clone_1b="Mitsui P1"
join(<1. 157,630. 706,796. 868,974. 1330)
/note="gene_id:MTH16.1
pir1|T05407
strong similarity to unknown protein"
/codon_start=2
/evidence=not_experimental
/protein_id="BAB17027.1"
/db_xref="GI:11022654"
/translation="1PFLLLAYFSRIYAGSLAENSHLIQSMATALFRSIRRDVVS
APLSVYKSLAGNAQPSWGSSYIGONVASFSAFGSKPVVNDIGTIGTNNAIREEER
KSKSTEAAIVGAQLTRSFALDVGTSKRLFTISGDIKTTQEEPKIKSFRLPSPLSY
YOPQNSMLSTIFNRISGVYLTGVTFAGYLLYLMGMICLTPSFYQVLYHTQOQLPVI
TSYTAALAIYHTIKSTHSLTH"
2019. 3572
/note="gene_id:MTH16.2
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09516.1"
/db_xref="GI:9758989"
/translation="MADDNCKNPNPNDRGIDTLTAFNNLKLYANYNHQEPGESGNTN
INSNNHNPPLITAFKSLKYANKNNOEPGESGNTINRNKNHNPPLVITAFNNLEL
YANYNHQEPGESGNTINNNNNPNPLITADNLRLYANYNHQEPGESNNNYPNLVY
NVGHISAASFNPPTPSSLTQPDSSSRYSQKPEPPPLSTFVSPGVDKDRMLSSL
DMTCAQRTFEFOXYLQDLDTYPTAERESHFKIGSALTTRKTRFLHLATNOYGSQAL
RILFRSPSIDHLIFCAVDITNFIILMSDKYGRLLIPAIRAVDKTKRESLYKLTGYET
LHLARLETGCLALNNVQEIIRGIYRDLIFECVANNADWLSFDPIGYTHVQNILIQNP
VATTAIAERLIRGSFRLAMEROGSYVEKCLSDFAHQVLEERGNAKEMVYRMITDK
FGNFVQSALRVMKEMRPLRFRFEVKLRPHFGKMEIGRCRNLRVYIOEEIVGWING
LPDKSGYVN"
complement(4620. 6215)
/note="gb|AA014519.1
gene_id:MTH16.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09517.1"
/db_xref="GI:9758990"
/translation="MDKFSNSYPDSAESSPRSRDVEFENPSWEDQOQNKYKIMC
SYGCKIQRPHDNQLTYNGDTKIMSVDGRIREFALVSKLSAVCSGGDGCIEISFKYQ
LPGEDLDALISVTNDEDLHMHMEYDRLLRLSTKPARMRLLFPSSPISGGFSGEST
KSDRDTLNPIPSRPESEKSVTAPNNADFLGSEKVAPIPPSPYKVPQVPPEPVLEP
PQMFVDORMLQPEHGVNPAEIQROIQEFQMIQIRDOEQQMLHQNQLHQQQQQQEAHQ
```

CDS

CDS

CDS

tRNA

CDS

```
NOLHQQEVITHONOLQOEALHQNOMLQOQOQOEAMYRKTEDEAGRYEPTVYQ
PVTHQOQPVGYWQGNNTNSNIQGNITYTTISQNLPEOQOQOQVYIMPAQSOAPGT
SVMRPIVOGNOGYPSPVQRLHHPDAYMEQOQOQVGNVQPOQPTFGSGQVMTSV
VMTSVGPRPMGLQEPYSQMKPVYTYTVAGGEMMYQPPPAQPOQOQOQOYQGMQPV
DLRGPDKGVAVNMAAPQVSDSV"
complement(6879. 8039)
/note="contains similarity to unknown protein
gb|AA015577.1
gene_id:MTH16.4"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09518.1"
/db_xref="GI:9758991"
/translation="MDVTGVTVRDAFDRVSKQKLYHSVTQVIDLYCDGIQDI
IQGNDQGVPEPSVLTFLRKRLDALLPIQLQSKHETKWSLSKYLKLEVSHP
LACFSVDFDINLVNKKILIHQCYREGFLDVGDQIVKAGREEIEVRSGQLEFHQI
LKRNIEPAMRWIFANRGLKQKSSKLEFKLSIKYCDILREGKSDALEYARTH
YPLHKEEIQKLTITCLWIGNFEKSPYAEIVSPSCWDKVTKELEIMEXHLLDQFIN
KVALSAGYESLPSLKLVLHMLALIKQEQWAMQCLPVPLELGNEYKFSAFVCPVS
SSEENPPMQLPQGHVVISQSMKRLSKNCAFRTPCPCPAETFLASACQQLYE"
5015,9236. 9366,5468. 9613,9767. 9843,9936. 10058,
-0157. 10259,10444. 10571,10659. 10745,10831. 10929
11073. 11152,11214. 11327,11415. 11531,11615. 11660
11749. 11829)
/note="gene_id:MTH16.5"
/codon_start=1
/evidence=not_experimental
/product="serine carboxypeptidase"
/protein_id="BAB09519.1"
/db_xref="GI:9758992"
/translation="MRNLSEFIVLFLTLTFPIHHLVDASLVLKSLPGEEGPLPFEI
YVSIQESGDVELFYFVFKSEKNPENDPLMWLTGPGSSICGLIFANGPLAFKGI
NGIYVPLELTSFSWTKYANILYLEAPAGSGSYAKTRAFESSDTROMHOIDQFL
EVKHPFEISNPFYVAGDSYQKIVPQAVQOISLVHSYIIEOGYVLGNPVTDKNIF
YRVPPAHGMGLISDELESLEBSGCGKFFNVDPNARCSNNLQATDHQMSFIYSEF
LRNCKVDYVLADTPNIRTDHRRTRYRYELSAFMANDENBRAIGVKKYPTGKNRCA
NIPYTFEIFNAVPIYHVNSLSKGRSLIYSGDHDSWVPFSSQAMIRALNTSIVDOK
WMMSSNQVAGYTRTYANKMTFATIKGGHTAETYPVQCSLMFRRWIDGEPPL"
complement(join(12194. 12286,12381. 12479,12565. 1274
12836. 12936,13097. 13257,13714. 13980))
/note="emb|CAB89365.1
gene_id:MTH16.6"
/codon_start=1
/evidence=not_experimental
/product="inorganic pyrophosphatase-like protein"
/protein_id="BAB09520.1"
/db_xref="GI:9758993"
/translation="MAATRVLTAAATAVTQTSGFLAKQAFTLPAAKSGCGFGGLCF
RALVYLSKRPFSQSAIYNPOVKQOEPAESLDYRVFFLDGSGKKVSPWHDIPITL
GVNFETIVEIPKESKAKMEVATDEDETPIKQDTKKGLRYYPYNINMNYGLPOTWE
SHAENSEVEGCFDNDPVDVVEIGETQKRGIDILKIKPLAALAMIDELELQKIVAI
DQPKHLVNDVEVDVEKHPGGLTATIRDMFRDVKIPDGKPAKRFGLGKPAKNDYAL
IQETNESWAKLVKRSVDAGDLSLY"
14190. 14266
/note="gene_id:MTH16.7"
/evidence=not_experimental
/product="tRNA-Val(AAC)"
complement(join(14508. 14580,14673. 14767,14841. 1493
15029. 15148,15236. 15325,15411. 15725,15814. 16032,
16352. 16414))
/note="emb|CAA10321.1
gene_id:MTH16.8"
/codon_start=1
/evidence=not_experimental
/product="microbody NAD-dependent malate dehydrogenase"
/protein_id="BAB09521.1"
/db_xref="GI:9758994"
/translation="MEFRGDANQRIARISAHILTPQMEAKNSVIGRENORAKGONPGI
VALIGAAGGIGOSLSILMKMNPVLSLHLHYVYNAQGVATADVSHMDGAVVRGFLG
QLIEDALTGMDLVIPAGIPAKPGMTRDDLPKINAGIVKTLCEGVAKCCPNAIVNLIS
PVNSTVPPIAAEVFKKACTYDPKLLGVTTLDVARANTFVAEVLGLDPREVDVPVWG
AGVTLLPQLSQVKPPSSFTPOEIEYLTNRINGGTVEVEAKAGAGSATLSMAYAAAK
ADACLRGLRGDANVVEGCSFVASQVTELAFATKVKRLGRTGAEVYQLGFLNEYERIC
```



CDS	EKADELASIOKGVFIRK" complement(16972..18612) /note="gb AAF24563.1 gene_id:MT16.9 strong similarity to unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB09522.1" /db_xref="GI:9758995"			
CDS	PNSHDEGTNIYDFMVEPERVHGGGLVMSLGGSIDRN" complement(join(20225..20505,20593..20908,21010..21060, 21130..21183,21341..21466)) /note="gene_id:MT16.10 unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB09523.1" /db_xref="GI:9758996" /translation="MILIRILDGQLDSLFRSLFSPLVFVSPILICLEQSQIHQTL LLPTFVCLDLRLLLPIKLKMDISRDDFTFSKVSPPDSEVKDLASPVGSITLKDGLD			
Query Match	4.58	Score 172	DB 8	Length 68098
Best Local Similarity	71.58	Pred. No. 5.5e-16		
Matches 226	Conservative 0	Mismatches 90	Indels 0	Gaps 0
QY 2970	ccgaaacctaattcaacttgagcgaagaagcgaagtacgatgaagacatgtagcga	3029		
Db 47712	CCGAGATCAAGAGACGACGAGAGAGAGAACTCTAGAGAACTCGGTGCTATGA	47653		
QY 3030	aggagatgcagtagatcgccgtcatgcaagccgtagacatcgaccctgcaagctcc	3089		
Db 47652	AGCAATGATGTACAGATCGCAGCCATCGCAATCGGTTGACATGACCCAGCAACCGTCA	47593		
QY 3090	ctaagccgaaccgcgaacgtaagcgaagcgaatcctcagaagctggtgtcgtc	3149		
Db 47592	AGAAACCCAAACGCCCTAACTGAGGATCTCCGACGACCTCAGAGTGTGGCGGTAGAC	47533		
QY 3150	ggcgtcgggaagatcagcgaagaatcctcaagagagatcgctgtgtgtgtgt	3209		
Db 47532	ATCGCCGTGAGAGATCAAGTGAAGATCAAAATCTTCAAGACTCGTGCCAGGTGGCA	47473		
QY 3210	cgaagatgacacagcttcacgtcgcagcgaagcacaacgttacaccaagtcttgaac	3269		
Db 47472	CTAAATGATACGGCTTCAATGCTCGATGAAGCTATACGCTATGTCAGTCTTGAAC	47413		
QY 3270	ggcaggtgagatctc 3285			
Db 47412	GGCAGATCCGGCTACT 47397			
RESULT 6	AT17114 92491 bp DNA	PLN	28-APR-2000	
LOCUS	AT17114			
DEFINITION	Arabidopsis thaliana DNA chromosome 5, BAC clone F17114 (ESSA			
ACCESSION	AL353994			
VERSION	AL353994.1	GI:7671408		
KEYWORDS				
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 92491)			

AUTHORS	Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A., Bancroft,I., Mewes,H.W., Rudd,S., Lemcke,K. and Mayer,K.E.X.
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 92491)
AUTHORS	EU Arabidopsis sequencing project.
TITLE	Direct Submission
JOURNAL	Submitted (28-APR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a>
FEATURES	Location/Qualifiers
Source	1..92491 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="5" 4369..5598 /gene="F17114_10" /number=1 4369..5598 /gene="F17114_10" /note="similarity to various predicted proteins, Arabidopsis thaliana" /codon_start=1 /product="putative protein" /protein_id="CAB89350.1" /db_xref="GI:7671409" /translation="MRSDLYTTVPCEFFKCPISLDVMKSPVSLSTGVTYDRVSIQ LDDGNNTCPATMOIIONKEFVNPILHLRIDHWSDSINRRADSESPDPTRDEI AIEFRRIENDARSKILRFARESDENREFLAGKDEVAMLVLDISDRNFSDSQLL EAVKILSMIRKRIEDRRRLSNLITNGDOLTFELLIKRGNPKLKIDCSAVLEFI DAESKLITAKEGVTEIITKLISDSDSLIEANLSLIAIASRVKRLALIREKI KLITSLTDPTSVSVTEKCLKLEAISGCKEGRSEICDGVCAETVYVNMKVSTAA HAYTVLWSVCYLYKEKKAQDAVIRINGVTKILLLLQSNCSLTVRHMLTDLKVFVR RCSLSYEITKTHIMP" 4369..5598 /gene="F17114_10" 7751..9517 /gene="F17114_20" join(7751..8105,8206..8286,8551..8780,8876..8976, 9063..9275,9358..9517) /gene="F17114_20" /note="similarity to Requiem protein, Xenopus laevis, EMBL:AB021738" /codon_start=1 /product="putative protein" /protein_id="CAB89351.1" /db_xref="GI:7671410" /translation="MATWYASSPAASPCSSNRRTKAPARRPSSPPPRKMKMAEI AKSVVVEQEEDEEDSYSNVTCCKGSGDEDEILICDKCDRGFMKCLRPVIVRV IGTWLVCDSQDQRYRRLSCKKLIHFRRIEKHTHOTCKLEISQETRRRRSCSLTV KRRRLPLVPSEDPDQRLAQMGTALGTYKSDGLNYPGMAPRSANOSKLF GGMQVLCKEDLETLEQCQSMYRRGECPLVVVYCPLEGYTVEADGPIKDLTFIAEYI DVDYLKNREKDCDSIMTLLISEDPSKIVICPDKGNISRFINGINNNPVAKKKU CKCVYSINGECRVLLVATRDISKGERLYDYNGYEHEPTTHFL" 7751..8105 /gene="F17114_20" /number=1 8106..9205 /gene="F17114_20" /number=1 8206..8286 /gene="F17114_20" /number=2 8287..8550 /gene="F17114_20" /number=2
exon	
intron	
exon	
intron	

exon	8551. .8780	/gene="F17114_20"	/number=3	8781. .8875	/gene="F17114_20"	/number=3	8876. .8976	/gene="F17114_20"	/number=4	8977. .9062	/gene="F17114_20"	/number=4	9063. .9275	/gene="F17114_20"	/number=5	9276. .9357	/gene="F17114_20"	/number=5	9358. .9517	/gene="F17114_20"	/number=6	10237. .11765	/gene="F17114_30"	join(10237. .10335,10761. .11055,11136. .11434,11532. .11765)	/gene="F17114_30"	/note="similarity to predicted proteins, Arabidopsis thaliana"	/codon_start=1	/product="putative protein"	/protein_id="CAB89352.1"	/db_xref="GI:7671411"	/translation="MASNPDLWCCKEERKQESFFKVLKRSDMSSSEDTRALPYDFVINFSDNELSGDMKFRVOWGNSWKVKISKNPRTFYEMKSGMEKEFVIDNALGDHEFLTTHKGQMSFTVKIFNKGDKEMQPPQSRASFASSRVKTEDVDYKREEVELVSSDSRSRGPITTA AETNRGSGYKRLKLNFGKKAEETQYKRTERTONSKRTERVSKERVYAGEPSSSVACFKIF:ISKSYIKSLAIPKPPGNYPKPKTRVKIHHPDGEKTKWVVFVVKERGQIFSGWKRLCKEYPVVEFGDTCKFTLLITPLELLLVSKP"	10237. .10335	/gene="F17114_30"	/number=1	10336. .10760	/gene="F17114_30"	/number=1	10761. .11055	/gene="F17114_30"	/number=2	11056. .11135	/gene="F17114_30"	/number=2	11136. .11434	/gene="F17114_30"	/number=3	11435. .11531	/gene="F17114_30"	/number=3	11532. .11765	/gene="F17114_30"	/number=4	12203. .13454	/gene="F17114_40"	12203. .12258	/gene="F17114_40"	/number=1	join(12203. .12258,12346. .12400,12637. .12687,12913. .13004. .13091. .13168,13316. .13454)	/gene="F17114_40"	/note="similarity to chloroplast ribosomal protein L17, Nicotiana tabacum, PIR:T01744"	/codon_start=1	/product="ribosomal protein-like"	/protein_id="CAB89353.1"	/db_xref="GI:7671412"	/translation="MTKFRKLRGPAGHRMSMLRTMVSQLVKNHERIFTTVTFKAKEVRRJ
------	-------------	-------------------	-----------	-------------	-------------------	-----------	-------------	-------------------	-----------	-------------	-------------------	-----------	-------------	-------------------	-----------	-------------	-------------------	-----------	-------------	-------------------	-----------	---------------	-------------------	---	-------------------	--	----------------	-----------------------------	--------------------------	-----------------------	---	---------------	-------------------	-----------	---------------	-------------------	-----------	---------------	-------------------	-----------	---------------	-------------------	-----------	---------------	-------------------	-----------	---------------	-------------------	-----------	---------------	-------------------	-----------	---------------	-------------------	---------------	-------------------	-----------	---	-------------------	--	----------------	-----------------------------------	--------------------------	-----------------------	--

	ADNM1QLGKEGSLSAARRACGFVRCDVYLHKIFTELAHRYKDRAGCYTRLLIRTRI
	DAAPMAVIFEENELROSKPATPQPPORVPILDPWERSRI.TROFAPPKEEKIPDSEI."
intron	12259.. .12345 /gene="F17114_40" /number=1 12346..12400 /gene="F17114_40" /number=2 12401..12636 /gene="F17114_40" /number=2 12637..12687 /gene="F17114_40" /number=3 12688..12912 /gene="F17114_40" /number=3 12913..13004 /gene="F17114_40" /number=4 13005..13090 /gene="F17114_40" /number=4 13091..13168 /gene="F17114_40" /number=5 13169..13315 /gene="F17114_40" /number=5 13316..13454 /gene="F17114_40"
exon	/number=6 complement(join(14357..15042,15306..16275)) /gene="F17114_50"
exon	complement(14357..15042) /gene="F17114_50"
gene	/number=1 14357..16275 /gene="F17114_50"
CDS	complement(join(14357..15042,15306..16275)) /gene="F17114_50"
	/note="similarity to pectin methyltransferase, Melandrium
Query Match	4.5%; Score 172; DB 8; Length 92491;
Best Local Similarity	71.5%; Pred.No. 5,3e-16;
Matches 226; Conservative	0; Mismatches 90; Indels 0; Gaps
OY 2970 ccgaaacctaatcacttgtagcaagaagacgaagatgacgtgaaacatacgatgacgcgatga	3029
DB 21872 CCGGAGATCAAGAAGACGACGACGAGCAGGAAGAACCTCTAGAGGAACTCCGTCTATGA	21931
OY 3030 aggaagatgcaatcatgatcgccgcatgcaagccgtagacacatcgaccctggcaagtc	3085
DB 21932 AGGAATGATGTACAAAGATCGCACCATGCATTCGGTTGACATCGACCACCAACCGTCA	21991
OY 3090 ctaagccgaacccgccgaacgaalaaagcaagatccctcagaacgltgtgtctgtc	3149
DB 21992 AGAACCACAACGCCGTAACGTGAGGATCTCCGACGACCCTCAGAGTGCTGGCGCTAGAC	22051
OY 3150 ggcgtcggyzaagatcacgagagaatccgaattctcaadagsgatgctgctgtgtg	3209
DB 22052 ATGCGCGTGAGAGAATCAGTAGAGGATCAGAATTCTCAGAGACTCGTCCAGCTGCGCA	22111
OY 3210 cgaagatgacacagcttccatgctcgaagacaacacatacqtacaccaagtcttgaac	3269
DB 22112 CTAAATGATGATACGGCTTCAATGCTCCGATGAAGCTATACGCTATGTCAAAGTTCTTGAAC	22171
OY 3270 ggcaagtgaggatctt 3285	
DB 22172 GGCAGATCCGCTACT 22187	



AC079733/c  
LOCUS 90448 bp DNA 19-JAN-2001  
DEFINITION Arabidopsis thaliana chromosome 1 BAC T8L23 genomic sequence,  
complete sequence.  
ACCESSION AC079733  
VERSION AC079733.1 GI:9994129  
KEYWORDS HTG.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 90448)  
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.C.,  
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,  
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.  
Arabidopsis thaliana chromosome 1 BAC T8L23 genomic sequence  
Unpublished  
2 (bases 1 to 90448)  
AUTHORS Town,C.D. and Kaul,S.  
TITLE Direct Submission  
REFERENCE Submitted (08-SEP-2000) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
3 (bases 1 to 90448)  
AUTHORS Town,C.D. and Kaul,S.  
TITLE Direct Submission  
REFERENCE Submitted (19-JAN-2001) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
COMMENT Address all correspondence to:at@tigr.org

BAC clone T8L23 is from Arabidopsis thaliana chromosome 1.  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Genscan+ (Chris Burge,  
http://CCR-081.mit.edu/GENSCAN.html), GenemarkHMM (Mark Borodovsky,  
http://genemark.biology.gatech.edu/Genemark/), Glimmer4 (a variant  
of Glimmer3, see Mihaela Pertea,  
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and  
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact  
mpertea@tigr.org), searches of the complete sequence against a  
peptide database and the plant EST database at TIGR  
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to  
indicate the level of evidence for their annotation. Genes with  
similarity to other proteins are named after the database hits.  
Genes without significant peptide similarity but with EST  
similarity are named as unknown proteins. Genes without protein  
or EST similarity, that are predicted by more than two gene  
prediction programs over most of their length are annotated as  
hypothetical proteins. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
Simple repeats are identified by repeatmasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES  
source  
1..90448  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="I"  
/map="mi259"  
/clone="T8L23"  
repeat\_region complement(634..656)  
mrna /rpt\_family="AT\_rich"  
complement(join(<1121..1284,1711..>1858))  
gene /gene="T8L23.1"  
complement(1121..1858)  
/gene="T8L23.1"  
/note="contains similarity to T box domain GI:5931592 from  
(Cynops pyrrhogaster)"  
complement(join(1121..1284,1711..1858))  
/gene="T8L23.1"  
/codon\_start=1  
/product="hypothetical protein"

/protein\_id="AAG50734.1"  
/db\_xref="GI:12321338"  
/translation="MALRQKLGWSEGLMRSDAKPCSRMLMKQTAALFTVGALGFK  
CRLHYGPRTVPRSLRWAGCGALSVSTAMLVRLFSPECEPQMIAYCNKIKTPQAP"  
complement(join(<3184..3267,3378..>3452))  
/gene="T8L23.2"  
complement(3184..3452)  
/gene="T8L23.2"  
/note="similar to salt stress induced hydrophobic peptide  
GI:1001779 from (Synecocystis sp.)"  
complement(join(3184..3267,3378..3452))  
/gene="T8L23.2"  
/codon\_start=1  
/product="salt-stress induced hydrophobic peptide,  
putative"  
/protein\_id="AAG50736.1"  
/db\_xref="GI:12321340"  
/translation="MGSLFLVLCALFIPVGVFLRYGIGLEFWCULLTLFAFIPG  
YAIYVLIK"  
complement(3452..3549)  
/rpt\_family="(A)n"  
complement(4220..4241)  
/rpt\_family="AT\_rich"  
complement(6494..6524)  
/rpt\_family="AT\_rich"  
join(<7494..7626,7715..7844,7921..8677)  
/gene="T8L23.3"  
7494..8677  
/gene="T8L23.3"  
/note="similar to DNA binding protein GI:19058 from  
(Hordeum vulgare)"  
join(7494..7626,7715..7844,7921..8602)  
/gene="T8L23.3"  
/codon\_start=1  
/product="DNA-binding protein, putative"  
/protein\_id="AAG50738.1"  
/db\_xref="GI:12321342"  
/translation="MKRHSQCYKOKLRKGLNSPEDEKLLNYTRKHGCMSSVPK.  
GLERCGKSCRLRWINTLRPDLKRGAFSSSEENLVELHAVLGNRMSOIAARLPGRIT  
EIKLWNSCIKKKLMKKGIIDPITHKPLSEVKEETNRSDNNSTSPSETNODLFVK4  
SDFAEYSAFOKEESNSVSLRNSLSMIPQFNIDGVSNSAGFDIOVCVKPSITL4  
PNTSSTVSGODHVNYSSEPWNSSGTSILNNPGMEEKWSFEYLNPSLFTQVY  
SETDFNSNIAFPWSQSQACDVFPKDLORMASFSGQT,,"  
complement(9315..9339)  
/rpt\_family="AT\_rich"  
<10278..>10856  
/gene="T8L23.24"  
10278..10856  
/gene="T8L23.24"  
/note="similar to hypothetical protein GI:9994090 from  
(Arabidopsis thaliana)"  
10278..10856  
/gene="T8L23.24"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAG50746.1"  
/db\_xref="GI:12321350"  
/translation="MKKPVYMKGGTLMRLNDGATVAFNPETFKARLISIRPKKEICS  
TILFTADNNILISATFEVYVAVFNILTDPKWVYLKQIRNGVLDKMLYSWYPEA  
DGKCLMLREITLKKDHYKQVLEGYD.RANKWEVIGSIPGWYTSALDFYOTPSLSSVI  
PDAKEEEETLACDHKKKISSINSIRMLDGISS"  
complement(11306..11348)  
/rpt\_family="(TA)n"  
complement(11470..11493)  
/rpt\_family="(TA)n"  
join(<12799..12984,13085..13522,13608..14039,14131..1  
14676..>14927)  
/gene="T8L23.4"  
12799..14927  
/gene="T8L23.4"  
/note="similar to jasmonate inducible protein GI:9279645  
from (Arabidopsis thaliana)"

CDS  
join(12799..12984,13085..13522,13608..14039,14131..14580,  
14676..14927)  
/gene="T8L23.4"  
/codon\_start=1  
/product="jasmonate inducible protein, putative"  
/protein\_id="AAG50741.1"  
/db\_xref="GI:12321345"  
/translation="MEVIGSTEGYTEDDSDHDDVFKIFVGGRQGIHY:FEFVKNQ  
QLESGVHLGVRYRGFTETFEINHLNNEHLEVEGYIDYGSYIOGLQFKTRVSELI  
GYDEGTRKFSLVKGRKRIIGFHGYNKERKIIISLGYPFSLPRKMEAKSGSGQWDDG  
TNNDGVTRKIHVRGVEGIIQYIKFDYVRKSGQHINGSIHGLSGSFTQIFEDHJNEH  
LVCVEGYDDESGVIALQOEKTNIKTSELGKKKFKSLVDKRRKIVGFHGAADKNL  
NSLGATTTVSPTKSECYGSSKGIYWDGVFDFIRTVVSSNMVMNRYIKFHYNRAY  
VVRQHGMSIVEEDGEKEFELDPNELTTSVEGTMSFSRSEIRISSLFKTSKGRTS  
PIGIASGTRKFLASKGCAVVGFGYGRHDDRDLVAIGAFESPLPPTAEKLGAGGONG  
DSWDDGVFEGRKRLYVGQKNCAFLKVYVDSNTQVVIQEDHGKNTLFEVKEYELEYR  
SEYITAVDGCYNKVNGETEVEVITMLRIQTNRKTSIPVGCESNSSFVLKKEGYKIVGFH  
GKASNMINDGVHVPLETE"  
complement(15377..15449)  
/rpt\_family="AT\_rich"  
join(<15722..16195,16329..>16895)  
/gene="T8L23.5"  
15722..16895  
/gene="T8L23.5"  
/note="predicted by genemark hm"  
join(15722..16195,16329..16895)  
/gene="T8L23.5"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAG50744.1"  
/db\_xref="GI:12321348"  
/translation="MDTLPHHLDELFRIDHRSIAMIOCTNRSIQSHISKPNPASE  
YFSIRSSLLIYATHGCPILCYHLYGDSRSRPTVETLQCHILGSCGLILFLMKNL  
CVANPLTKKFRFLNHSRKFPTDFTLRGRQQVIGFADQTDRTTPSFKIVNIKEGA  
VHMLRNDGSIYAFNLKTEKARLPIRPEKELCLTFEAVDNLLTISATEEYIYVYA  
FHNILSDPKWVLYKQIRNGVLDERLWSYVEAYDGKSLVLEETILKKDYNNKHLQV  
LHGVDLRANKWEVIGSIPWSPGLDFYQFTSPSSVIGLDDKEGEELIACDDKRIPS  
VNAILRLIYGIS"  
complement(join(<18101..18211,18372..18517,18621..18726,  
18800..18877,18965..19018,19096..19156,19248..19459,  
19540..19613,19700..19829,19912..19994,20071..>20269))  
/gene="T8L23.6"  
complement(18101..20269)  
/gene="T8L23.6"  
/note="similar to pectinacetyltransferase precursor  
GI:1431629 from (Vigna radiata)"  
complement(join(18101..18211,18372..18517,18621..18726,  
18800..18877,18965..19018,19096..19156,19248..19459,  
19540..19613,19700..19829,19912..19994,20071..>20269))  
/gene="T8L23.6"  
/codon\_start=1  
/product="pectinacetyltransferase precursor, putative"  
/protein\_id="AAG50747.1"  
/db\_xref="GI:12321351"  
/translation="MKKLMSWILIFNIHVNGMMMEFDENEMFTVFNGTKVQTONDV  
Query Match 3.5%; Score 133.2; DB 8; Length 90448;  
Best Local Similarity 54.8%; Pred. No. 2, 2e-10;  
Matches 311; Conservative 0; Mismatches 248; Indels 9; Gaps 2;

DB 65452 CTGTGCAAGTGTCTTCGGAGGTGTCCCTCCTCCCTACACTCGGAGATGMAAGTTTA 65393  
QY 1640 atgagqcccatgcaactatgatgtcaacglttattttcacaataactatatata 1699  
DB 65392 ATTCGGGTATGAGAATAATCCTTCAAGAAGCATTGATGTGAACCTTTTGAAACAGAT 65333  
QY 1700 ttcaaaaataagtttgcagataaaattttttgg--tgytcaagctttctccaact 1757  
DB 65332 TGTAGAGATTAGTCTTCGATGTGTCAGTCTCTTGAAGATGACCACTTATCCAATCT 65273  
QY 1758 actaaactagtttggaaatgtctctcttcttatttttcttctctgattcttatgttt 1817  
DB 65272 ACTAGATGAGTTAGAAATGCTTGTGTGTTAATTTCCGGTTCGCAATTTCTTATATTTC 65213  
QY 1818 ttattatggaattttaaagacgactgtttaagtcgttctctctctctctctgttctctaa 1877  
DB 65212 TCGTTTATATAATGTGAAGCTGATGTGTTAGTTCGCTCTTCGTTTACTGCTTTCGGA 65153  
QY 1878 agttactttgttaactcatctcctcccaattagacagtcacatcatatagttatcttta 1937  
DB 65152 AATTATTTTATAGACTATTTCTCTCCAAATTAGGCAATCAATCATGAAGTTATATTTTA 65093  
QY 1938 atatatgtcagtcgataaaaaaatga 1965  
DB 65092 AATAATTTGTGGTTGATTAATAAAAAAAGA 65065  
RESULT 8  
AC079730/c standard; DNA; P.N: 90448 BP.  
ID AC079730  
XX AC079730;  
AC AC079730.1  
SV AC079730.1  
XX 12-SEP-2000 (Rel. 65, Created)  
DT 12-SEP-2000 (Rel. 65, Last updated, Version 1)  
XX Arabidopsis thaliana chromosome 1 BAC T8L23 genomic sequence, complete  
DE sequence.  
XX HTG.  
XX Arabidopsis thaliana (thale cress)  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
XX {1}  
FN 1-90448  
RP 1-90448  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,  
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.  
RA Bowman C.B., White O., Niernan W.C., Fraser C.M.;  
RT "Arabidopsis thaliana chromosome 1 BAC T8L23 genomic sequence";  
RL Unpublished.  
XX {2}  
FN 1-90448  
RP 1-90448  
RA Town C.D., Kaul S.;  
RL Submitted (08-SEP-2000) to the EMBL/GenBank/DBJ databases.  
RL The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD  
RL 20850, USA, cdtown@tigr.org  
XX SPTREMBL: Q9FDZ9; Q9FDZ9.  
DR SPTREMBL: Q9FE70; Q9FE70.  
DR SPTREMBL: Q9FVS7; Q9FVS7.  
DR SPTREMBL: Q9FVS8; Q9FVS8.  
DR SPTREMBL: Q9FVS9; Q9FVS9.  
DR SPTREMBL: Q9FV70; Q9FV70.  
DR SPTREMBL: Q9FV71; Q9FV71.  
DR SPTREMBL: Q9FV73; Q9FV73.

DR SPTREMBL: Q9FVT4; Q9FVT4.  
DR SPTREMBL: Q9FVT5; Q9FVT5.  
DR SPTREMBL: Q9FVT6; Q9FVT6.  
DR SPTREMBL: Q9FVT7; Q9FVT7.  
DR SPTREMBL: Q9FVT8; Q9FVT8.  
DR SPTREMBL: Q9FVT9; Q9FVT9.  
DR SPTREMBL: Q9FVT0; Q9FVT0.  
DR SPTREMBL: Q9FVU1; Q9FVU1.  
DR SPTREMBL: Q9FVU2; Q9FVU2.  
DR SPTREMBL: Q9FVU3; Q9FVU3.  
DR SPTREMBL: Q9FVU4; Q9FVU4.  
DR SPTREMBL: Q9FVU5; Q9FVU5.  
DR SPTREMBL: Q9FVU6; Q9FVU6.  
DR SPTREMBL: Q9FVU7; Q9FVU7.  
DR SWISS-PROT: Q9FVT2; EFlH\_ARATH.

XX Address all correspondence to:atctigr.org  
CC BAC clone T8L23 is from Arabidopsis thaliana chromosome 1  
CC The orientation of the sequence is from Sp6 to T7 end of the BAC clone.  
CC Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GenemarkHMM (Mark Borodovsky, http://genemark.biology.gatech.edu/Genemark/), Glimmer4 (a variant of GlimmerM, see Mihaela Perlea, http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and Genesplicer (Mihaela Perlea and Steven Salzberg, contact mperlea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SF/). Simple repeats are identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Key Location/Qualifiers  
FH  
FH source 1. .90448  
FT /chromosome="1"  
FT /db\_xref="taxon:3702"  
FT /organism="Arabidopsis thaliana"  
FT /cultivar="Columbia"  
FT /map="m1259"  
FT /clone="T8L23"  
FT repeat\_region complement(634. .656)  
FT /rpt\_family="AT\_r1ch"  
FT mRNA join(complement(1711. .>1858),complement(<1121. .1284))  
FT /gene="T8L23.1"  
FT join(complement(1711. .1858),complement(1121. .1284))  
FT /codon\_start=1  
FT /db\_xref="SPTREMBL:Q9FVU7"  
FT /gene="T8L23.1"  
FT /product="hypothetical protein; 1858-1121"  
FT /protein\_id="AAG12106.1"  
FT /translation="MALRQKLGSSEGLMRSDAKPCSRIMKQTAIFTVGALGFVWLCRLHYGRITVPRSLRWAGCGALSVSTAMLVRLFSPECEQNIAYDNIKTPQASLP"  
FT mRNA join(complement(3378. .>3452),complement(<3184. .3267))  
FT /gene="T8L23.2"  
FT join(complement(3378. .3452),complement(3184. .3267))  
FT /codon\_start=1  
FT /db\_xref="SPTREMBL:Q9FE70"  
FT /gene="T8L23.2"  
FT /product="hypothetical protein; 3452-3184"  
FT /protein\_id="AAG12109.1"  
FT /translation="MGSFLEVICAIFIPVGVFLRYGLGLFVWVCLLLTLFAPIFGLIYAIYVLTK"  
FT CDS join(complement(3452. .3549)  
FT repeat\_region complement(3452. .3549)

FT /rpt\_family="(A)n"  
FT complement(4220. .4241)  
FT /rpt\_family="AT\_r1ch"  
FT complement(6494. .6524)  
FT /rpt\_family="AT\_r1ch"  
FT join(<7494. .7626,7715. .7844,7921. .8588,10051. .>10856  
FT /gene="T8L23.3"  
FT join(7494. .7626,7715. .7844,7921. .8588,10051. .10856)  
FT /codon\_start=1  
FT /db\_xref="SPTREMBL:Q9FVU6"  
FT /gene="T8L23.3"  
FT /product="hypothetical protein; 7494-10856"  
FT /protein\_id="AAG12112.1"  
FT /translation="MKRHGCCYKQKLRLKGLWSPFEEDKLLNYITKHGCGWSSVPLERCGKSCRLRWINYLRPDLKRGAFSSSEONLIVELHAVLGNRWSQIAARLPGRITKNIWNSCIKKKLMKKGIDPITHKPLSEVGKETNRSDNNNSTFSSETNQDLFVAKTAEYSAFOKEESNSVSLRSLSSMIPTOFINDDGSVSNAGFDQVCVKPSIILPFPSTSVSGODHVNVSSEPWNESGCTSHLNPMEEMKWESEYLNESJFTQVYVKSINSKIIVINEVRNSNETYEFEINVGYSWKLSE7TLTCTSNLDDRMKPPVYMKGLRNDGAIIVAFNPETEKARISIRPKELCSKTLFTADNNLILISATEEVYVAVANLTDPKWVVLKQIRNGVLDEKMLYSWPEAYDCKLMLREILKKDHYKQVLHGTYDLMEVYGISPGWYTSALDFYQFTPSLSSVIGPDAKEEELIACDHKKISSINSIIRMLSS"  
FT complement(9315. .9339)  
FT /rpt\_family="AT\_r1ch"  
FT complement(11306. .11348)  
FT /rpt\_family="(TA)n"  
FT complement(11470. .11493)  
FT /rpt\_family="(TA)n"  
FT join(<12799. .12984,13085. .13522,13608. .14039,14131. .14676. .>14927)  
FT /gene="T8L23.4"  
FT join(12799. .12984,13085. .13522,13608. .14039,14131. .14676. .14927)  
FT /codon\_start=1  
FT /db\_xref="SPTREMBL:Q9FVU5"  
FT /gene="T8L23.4"  
FT /product="hypothetical protein; 12799-14927"  
FT /protein\_id="AAG12115.1"  
FT /translation="MEVIGSTEGYTFDDGSDHDQVTKIFVGGRGQIHVIEFEVVKLESQVHLAGVRYRGFTIEFEINLHNEHLESEVEGYDYGSYITQGLQKTFNRYSETDEGTFKLSLVKVGKRIIGFHGYMKERKILSLGYSFIHPRKMEAKSGKGNQNDQDGVTKIHVRKGVFCIOYIKFDYVRKSGOHINGSIHGSGSGFTQIFEIDLNDENHEGYVDDDESGVIALQFKTNIKTSELLGYKKKKFSLVDKRRKIVGFHGYADKNNLSYFTVSPTKSECYGSSKGIYWDGVEDFIRIVYVSSNVMNVRIKFHYNNRAVYVFWNSIVAEDEGEKEFELDYPNELLITVEGTMKSPSRSEIRISLITFKTSKGRTPYICGTKFLASKGCAVVGFGYGRHDDRDLAIGAYFSPLEPPTAEKIQAOGNGDSDMDCEGVRKLYVGQGKNCVAFLLKVYVDSNTQVIGEDHGKNTLFEVKEYELEYPSEYIIIAQYNKVNCTEEVITMLRIQTNRKTSIPVGCESNSSVLLKKEGYKIVGFHKGASNNLGVHVPLTE"  
FT complement(15377. .15449)  
FT /rpt\_family="AT\_r1ch"  
FT join(<15722. .16195,16329. .>16895)  
FT /gene="T8L23.5"  
FT join(15722. .16195,16329. .16895)  
FT /codon\_start=1  
FT /db\_xref="SPTREMBL:Q9FVU4"  
FT /gene="T8L23.5"  
FT /product="hypothetical protein; 15722-16895"  
FT /protein\_id="AAG12118.1"  
FT /translation="MDTLPHLJLDEILFRIDHRSIAMIQCTNRSLOSHISKTPNFNFSRIRSSLIYIATHGCPPLICVHLGDSRSRPVEYETLTQCHILGYCSGLLLLELNKNLNDLTKEFRFLNHSRSKFFPTDFTLRGGRQVIGFADQTDRTTPSFKIYVNIKEGALRNDGSIIVAFNLKTEKARLPIRPFKELCKTLFTAVDNNLTLLISATEEVYVYAEALSDPKWLVKQIRNGVLDKRLYSWYVEAYDKSLVRETIILKKDYNNKLCQVLAIRANKWEVIGSIPBWSPSGLDFYQFTPPSSVIGLDDKEGEEIILACDDKR1PISVNALIYIGIS"  
FT join(complement(20071. .>20269),complement(19912. .19994  
FT complement(19700. .19829),complement(19540. .19613),  
FT complement(19248. .19459),complement(19096. .19156),  
FT complement(18965. .19018),complement(18800. .18877),  
FT complement(18621. .18726),complement(18372. .18517),









/product-"hypothetical protein"  
/protein\_id-"AAC71888.1"  
/db\_xref-"GI:3845199"  
/translation-"MFALKKNTVREGFVNICFSYLKLLYKSNFVTNLNETNNEKR  
TNKRIYKKSQAOSLFQKGLNIHDKLLEFKNLPKYKCAKECISAKVEYKYLLDEYKCC  
FNYISLCDIIQSVKIEDELDKTFEDYNEVIEVKNIQKNVLRKINEIYFKNKDITFHR  
EILGKICNKIMSYIHEMGNELIHLIYEFPRMKNKDKNLIFPNYFNVFDMHMLFN  
HEIYKLLFIENKYLNNSNIPEFNKLIQEMFNLYFRELKNEKNYIIKMKKSEIYKK  
CFAKFHENVDHIQNEKILNIRLYVDSILDIDINMKMLNMLNINENIEYISKLL  
NFYCTLLIKGKYDNMTIYKLEKVIKATHHILCDKTKNTEFCSDIDYSTLSLNK  
FILNKIQRNEFLFEYECCLKILNIRKFNQSLCISLSKNTYNNILRMNVYVNV  
LFNDIMKFSYLCNIFLGKRKIKTEENAVLIHNNQNTNKNENIKDIIQKRIKEY  
IFYKMENYKDFHFKLKDSLSIKLSNTFVKINEVNTSYDFYLLFNINISCIYNFY  
NRNSVKKKKDYIYILNDLSFYVKYIKNNDRTKKKNFLLSSMKELICKNLSVSN  
RYIKHLHEEDNFQKQYVCSLTFELNLFEDKIHFHYIYLMCHVYKTYNFKCNKL  
INEDIISLLLTCSKFOYFIENNSNDRCRKLHLKYNIDDLIKNYLNTYKSID  
NISKIFISLSNSKYTCENVENLLESQSEFEKVTKTSKGGIHMMDNLLDNNNSCE  
KYEHRVIEYKKENLFINLKIIECLIKNIFLYLKKKKTYLYLKQSLCPINLKENIL  
KKILYIANNLMYEMYGYVCEMLERYLSHKEQNLFSYNNKKNVEHKMFQDKILCHISE  
DOYIEMSNTRYVLYFDYLNKINSEOSNILRNSTNDRFDEIKEKKYKLNNTLIK  
NNVKLNYESKSNNSMGNISNLIKDKRKNHNNVEMDLIDNKNENKRIQEKQNGENSEN  
CKDLVNDIINIFGFLMEKKKFLFEQOLMYLCNITKFKRRYVSSSLFHMVFEKIIK  
DMNLKYLCLENYKIKNECAFLYIIDIVLFEKER"

BASE COUNT 6284 a 1019 c 1106 g 6458 t  
ORIGIN

Query Match 3.3%; Score 128.4; DB 3; Length 14867;  
Best Local Similarity 43.7%; Pred. No. 1.2e-09;  
Matches 1172; Conservative 0; Mismatches 1466; Indels 44; Gaps 12;

QY	116	aatttcgtaaaagaaaaacctatgtaataatcacgcgtagcgcaaaaaataatcctatc	175
Db	9666	AAATTATTAATTTATTTATTAATAATTCATATATCATTTATTTCAATTGCAAGTTGCTAA	9607
QY	176	ataataaatttlatatagaaltatacattctctcgtaagaagttactccaattgca	235
Db	9606	ATTTTATATATTATTATTA--TAAATATATATATTATTTAATATATAAATTAATTGTA	9550
QY	236	agtggtgalttaactaataaagaagcaagaagtagaagcttataatlaattgattgctc	295
Db	9549	TTAAATTAATAAAAAAAATATATATTATTGTTAAATTAATATACATTATATATATTTTAA	9490
QY	296	atagtaatggtatattggtgtaataatacaagtaacatcaatcaatgcatgaacgaggg	355
Db	9489	TTTTAGTATTAAATTATTATATATATTAAAAAGAAATATATTTTTTTATATAAATAT	9430
QY	356	tatttgcttgtaactcatgaagcgcaatgtaagaagatgtgaagtcctcatattgaa	415
Db	9429	AAATGCTGTAATATATAAATATTAAGATTAAATTAATAAATTAATAATATTTTAA	9370
QY	416	aatttaccctctagcttctgcgatttaactctatgaatgaacgcaacatatagaatt	475
Db	9369	AATTAAATATATAATAATATTATAAATAATATTATAATTAATAATAATTAAT	9310
QY	476	tcatgtgcacaacgacatttagacggtatcttcaattagaccgattcaattagatatact	535
Db	9309	TAAATTAATAATAATAATTTAAATATTAAATTAATTTAAATTAATTAATTAATAA	9250
QY	536	tatatataaattagtggtgataaagtttactctaccacttgagaacttacaacaatgg	595
Db	9249	TTAAAAAATTAATATATAATAATAATAATAATTAATTTAAAAATTAATAATA--	9192
QY	596	gcaataaccttaatgtcgaagaagccgctcccaactcggtgaatgagttatgggggagag	655
Db	9193	-----AACTAAATTAATAACAATATAATAATAATAATAATAATAATAATAAT	9139
QY	656	atcctgttaaatcgtaacaataaacttaagaactagaatgacaccaaaaaatacata	715
Db	9138	AATATTAATAATAATAATTTAATTAATAATATAAATAAATTAATTAATAATAAT	9079
QY	716	aaagaaagcttgaaagaagtcattatcgtatccagctcatatttccagtaagaatcaaa	775

Db	9078	AAATTAATAATAATTAATTAATTAATAATAATAATAATAATAATAATAATAATAATA	9021
QY	776	tcaagcgcttgaaagcgcttgtaagaanaatgtaagaagaacgcgtggggtttagaagaa	835
Db	9020	TAAATTAATAATTAATTAATTAATAATAATAATAATAATAATAATAATAATAATA	8961
QY	836	gacaagaataagaagaacaatgattgtaattgctctatttgggtgatalaggagttgcaa	895
Db	8960	AATATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	8901
QY	896	aaagagagagagagaagaagaattaggtcaaaaataatgaagcactaaaaatgagagacatqg	955
Db	8900	AAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	8843
QY	956	tgaagtaactatcacagaagcgacttatgcttctcttatggcaatgataatccaacccaag	1015
Db	8844	-----AATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	8791
QY	1016	tgaacgctcccttttggccctaatctcgtaagctctctctctctcgtcccttaggaa	1075
Db	8790	ATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	8731
QY	1076	aaaccctagaagaatttaacccttgctctgatactgctctttagtaaacatgatttga	1135
Db	8730	AAATTAATTAATGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8671
QY	1136	ccacaactattctctctatcttcttggtgctataaggaatttgcctlatatgcttctt	1195
Db	8670	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8617
QY	1196	gtattgctccgtacgtacgatacgaattlaaatgtrtataacaaggttataataaacta	1255
Db	8616	TAAATTAATGAATTAAGCAATATTATTATTATAAAAAATAAATTAAATTAATTAACCTT	8557
QY	1256	gcacaatlgatccatgaatcttgtagcgaagaagtagaataatataatgagctttaa	1315
Db	8556	TAAATTAATACATATTATTATTATTATTGGAATTAATTAATTAATTAATTAATTAAT	8496
QY	1316	cggcaatatataaatttgcgtgcaaaactagcttcaatcatgatactaatgatacttctc	1375
Db	8497	----AATTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8442
QY	1376	ttaatttcccttggccaatatacatgacatgcacggaatttggcgaagttatgctgcgaat	1435
Db	8441	TAAATTAATAATAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8382
QY	1436	tcttccatcac-acaaacactaaacttaattagaactcttaggaaatattttaaatgaca	1494
Db	8381	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8322
QY	1495	acttatacgaaaaaaattagtaatgaaacaactccagaatlaaacatgagctatataa	1554
Db	8321	TTAAAAAATAATTTAAATATATATTTAATTTAAATTAATTAATTAATTAATTAAT	8262
QY	1555	ctlaagataaaatgaagtaataatgataatgataataacatalcgtatgaggttaaaa	1614
Db	8261	ATATATATATCATTAATTTAATAT--ATTTTTTATATAATAATAATAATAATAATA	8204
QY	1615	aaaaacatactgataaattgctacatgcagcccatgcaactatgataatgataacggtt	1674
Db	8203	TATTAATAATAATAATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTT	8144
QY	1675	ttatttcacaataactaataatataatca--aaaaatgcttctcagatlaaatctt	1732
Db	8143	AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8084
QY	1733	tttggtgctcagcttctcccaacctaaactgaatggaatgcttctctcttattt	1792
Db	8083	ATAATTAATTAATTTTATATGTAATTAATTAATTAATTAATTAATTAATTTTAAAT	8024
QY	1793	tcttctctgacttctlatgcttcttatttataatgaactttaaagcggatgcttaggtc	1852
Db	8023	AAATTAATTTATATTATTATTATTATTATTATTAAACATTTAAAAAATAAATAATTAATTA	7964



Db	13782	ATACATACATGATATTTTAAATAATATATATAATCCATATGTTGTTATATTATATATAT	13841
Qy	2051	ggtlaaatttagtaatttttacaataaagaactgtaattgtagtataattgaggaa	2110
Db	13842	ATATATATATATTAATAATATTTTATTTTATTTTATTTTAAATTTAAATAATAAA	13901
Qy	2111	tcaagtataattagcttggttaactatactctcggaacttgaagatttaactattgca	2170
Db	13902	TAAGAAAAAGAAAAATATAATATATATATATATATATCTTTTAAAAAAATATTTT	13961
Qy	2171	aaattataaatttagtccccgaaaaatacagacgacgagcaacacataaagcaggt	2230
Db	13962	TAAATAAAAAATTATTAATAAAAAATAATATATATTTTAAAAAAATTTTAAAGTTT	14021
Qy	2231	tgaatcttggaataatttgcatacataacctataaactaactaactgltctggtggttc	2290
Db	14022	AGGTTTGTAGTAGTTTAAATTTTAAAAATAAAAAATATAATTTTCTTTTCTTTT	14081
Qy	2291	aaaagcccttccaagaagttccatttttaaatccaagacatttacaagaataaagt	2350
Db	14082	ATTATTTTATTTACAAATTTGTTTATTTCTAGTTTAAATTAACAAAAAATAATTT	14141
Qy	2351	tgaatcaaaaaataaatggtattttgtaaggtttttttgtaaaaaagcacat	2410
Db	14142	ATATATTTTATATTTTAAATTTTATATTTAAATTTTAAAGTTTAAAGCATTTCTATTTAT	14201
Qy	2411	taagaagttagtttttttccacccaataatcaatlaatttaaaccatgcaaccattc	2470
Db	14202	ATATAAAAAAGAATATATATAATATAATATATATATATATATATATATATATTA	14261
Qy	2471	ataaacaataactatlaaagaataataatcacaaaataattaaatacacctlaaaatt	2530
Db	14262	ATACAACCATATTTAT-----ATATAAACAATATTTATATATATATATATAGATTCAACT	14315
Qy	2531	acataataatttacaacaacatctaattgaacacagaaaggaaggaataatatact	2590
Db	14316	GTAGTATATATATATATATTTAAAGTTTAAATTAATATATATATATATATATATATAT	14375
Qy	2591	ataaatgagacatatatacctataaataaataaatgagcatalgaaataataaagag	2650
Db	14376	ATATATATATATATATATATATATTTTATTTTATTTATATATATATCTAAGTAATAAT	14435
Qy	2651	acatgcatgtaagcattcggttaattaatcgagtcagaatatataatcagtaatacata	2710
Db	14436	AAAAGTAATAAAAATATATATTTTATTTACATCTGTCAACAAATCCTTTTAGATATATTA	14495
Qy	2711	tgtgtataattctggaataaataataataatgagaataaagaataaagatgaataatgaa	2770
Db	14496	TTTCTGTAGTATTAATCTGTATATATATTTTAAATATATATATAGATATAAATAAT	14555
Qy	2771	aatgtaataataaagaaga 2791	
Db	14556	AAAAAAAAATATATATATATA 14576	
RESULT	12		
LOCUS	DMU11584	4601 bp	DNA
DEFINITION	Drosophila melanogaster Oregon-R mitochondrial A-T region.	INV	23-JUL-1994
ACCESSION	U11584		
VERSION	U11584.1	GI:508826	
KEYWORDS	mitochondrial DNA; A+T region; tandem repeats.		
SOURCE	fruit fly.		
ORGANISM	Mitochondrion Drosophila melanogaster		
REFERENCE	1 (bases 1 to 4601)		
AUTHORS	Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.		
TITLE	Sequence, Organization and Evolution of the A+T Region of Drosophila melanogaster Mitochondrial DNA		
JOURNAL	Mol. Biol. Evol. 11, 523-538 (1994)		
MEDLINE	94285822		

REFERENCE	2 (bases 1 to 4601)
AUTHORS	Kaguni, L.S.
TITLE	Direct Submission
JOURNAL	Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of Biochemistry, Michigan State University, East Lansing, MI, 48824-1318, USA
FEATURES	
Source	Location/Qualifiers
	1..4601
	/organism="Drosophila melanogaster"
	/organelle="mitochondrion"
	/strain="Oregon-R"
	/db_xref="taxon:7227"
	/dev_stage="embryo"
	650..1022
repeat_unit	/note="repeat I-A"
	/rpt_type=tandem
repeat_unit	1023..1360
	/note="repeat I-B1"
	/rpt_type=tandem
repeat_unit	1361..1705
	/note="repeat I-C/A"
	/rpt_type=tandem
repeat_unit	1706..2043
	/note="repeat I-B2"
	/rpt_type=tandem
repeat_unit	2044..2388
	/note="repeat I-C"
	/rpt_type=tandem
misc_feature	2491..2511
	/note="deoxythymidylate stretch"
repeat_unit	2512..2648
	/partial
	/rpt_type=tandem
repeat_unit	2649..3112
	/note="repeat II-A"
	/rpt_type=tandem
repeat_unit	3113..3576
	/note="repeat II-B1"
	/rpt_type=tandem
repeat_unit	3577..4040
	/note="repeat II-B2"
	/rpt_type=tandem
repeat_unit	4041..4504
	/note="repeat II-C"
	/rpt_type=tandem
misc_feature	complement(4565..4585)
	/note="deoxythymidylate stretch"
BASE COUNT	2271 a 131 c 74 g 2125 t
ORIGIN	
Query Match	3.2%; Score 125.2; DB 3; Length 4601;
Best Local Similarity	44.4%; Pred. No. 3.9e-09;
Matches 1187; Conservative	0; Mismatches 1438; Indels 49; Gaps
Qy	99 ttgaactgaacattctaatttggtaaaagaaaaaacctatglttaatacacccgtagc 158
Db	312 TTTTTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAA 371
Qy	159 aaaaaatactcatcataaatttataataagattatacatctcgttgtaa 218
Db	372 AAATATAGATTAAATTTCTTTAAATGACTAAAAAAATTTTAAATTTTAAAGTATTTTAA 431
Qy	219 gaattactcaattgcaagtggtgtaataaataaagaagcaagtaggaagctta 278
Db	432 ACTTTT---TTACAAATTTTAAATAATATATAATTAATTTTAAATAATTTT 488
Qy	279 taatlaatgatgltgcataglactglatatgltgatgaataaagaagtagaagcat 338
Db	489 TTTTAAAAAATAATGAAATATATATAAATAATTTTAAATAATGAAATAATTT 548
Qy	339 taatgcatgaacggggtattttgcttgaaactcatlaaaggcaatgtgaagaagatg 398

Db 549 AATCTATTAAAAAAATTATTAAAAATTTTATGCAATAAATAAAAAAAGTAAATAATTTA 608  
QY 399 tgaagctcattctgaaaatttattctttagcttctgcatltaaatctatgaaatgaa 458  
Db 609 TTAAAAATCAATATATATATATATAATAATAATTGATATATAATTAATTAACGAATAA 668  
QY 459 cgcaacatatagaaaatttcattgtgacaaacgacatttagacggtatcttaattagaccga 518  
Db 669 TAAATATATAATAATAATTATTTAATCACTAAATCTGAAATATAATTATATATATATAT 728  
QY 519 ttaattagtaataacttatatatataataattagtggtgattcataagtttactatccactt 578  
Db 729 ATATATATATATATACATATATATATATA-----TATACATAAATAAATAAAT 780  
QY 579 gagaatttaaaccaatgaggcaataacccttaatgctgaaagaagccgctcccaactcgctga 638  
Db 781 AAATTTATCCCCCTATTCATAAATTTTATATATAATTAAACTT---AAAAAGTATTT 836  
QY 639 tgaattatggggagagagatccctgtaaatcgtaacaataaacaacttaagaactagaat 698  
Db 837 TTTTAAAAAAAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 896  
QY 699 tgacacccaaaaatcatataaagagaacgctggaagaagtcatttaccgtatccagctcatat 758  
Db 897 TTTTATAAATAAATAATTATTAAAAATTAATTAAGAAATATTTTATATATAATTAATAATTT 956  
QY 759 tcctagctaagaatcaaatcaagccgctggaagggctgtlaagaaaatgctggaagaacc 818  
Db 957 AA--AAATAATTTTAAAAAAATTCAAATTTTATTTATATATATATATATATATATATATTT 1014  
QY 819 gtggggtctagaagaagaagacaagaatagaagaacaatgattgtaaatctgctatcttg 878  
Db 1015 TAATTTCAATTAATTTATATATAATATAATAATAATAATAATTTTAACTCACTMAATCTGA 1074  
QY 879 tglataggagtgtaacaagaagagagagaagaagaataatagtcacaataatgagcaact 938  
Db 1075 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTCCCCCT 1134  
QY 939 aaaaatggagacatgcttgtagtaactatcaagaagcgactatgcttccctatgacca 998  
Db 1135 ATTCATAAATTTATTTGTATTAATTAATAAACTTAAAAAATATTTTTTTAAAAAAAATGA 1194  
QY 999 tgalatccaacaagaagtgcaagctcctcttcttgccctaatttgcgtaaagctcctcc 1058  
Db 1195 T-TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1253  
QY 1059 tctctgcctctaggaaaaaaccctagaatcttaacccttgctcctgactctgctcttg 1118  
Db 1254 TTTAAATTAATTAATTAATAAATAATTTTAAATTAATAAATAAATTTTAAATTAATAA 1313  
QY 1119 agtaaccatgatttgaccacacactattctctctatcttcttggtctatagatttg 1178  
Db 1314 AAATTCATTCATATATATATATATATATATATATATATTAATTTCAATTAATAATTAAT 1373  
QY 1179 cttlatatgcttcttgatctgctccgtaagctatagcgaatttaaatggtlataac 1238  
Db 1374 AAGTATAATAAAAATAATTATT---TTAATCACTAAATCTGAATTAATTAATTTGTATA 1428  
QY 1239 aaggtctatataaactagcacaatgagtcocatgaaatttgctlagcgaaaaaggtagaaa 1298  
Db 1429 TATATATATATATATATATATGTAATAATGAATAATAAATTATTCCTCAATTAATTAAT 1488  
QY 1299 tatattgagctctttaaacggcaatatatatataatttgcgcgaanaacttagctttaatcat 1358  
Db 1489 TATTATATTAATTAATAACTTAATAAAAAATATTTTAAAAAAAATAATTAATTAATTAAT 1548  
QY 1359 gatctaataatattcttctaatttcccttgcccaaattaatacacatgacggaatttctg 1418  
Db 1549 ATACTTAATTAACCTATTTTATTAATTAATTAATTTTATTAATTAATAAT---TATTTAAAA 1603  
QY 1419 caagttatgctgcgaattcttccattcacacaacaactaaacttaataagaactctagaaa 1478  
Db 1604 TAAITTAATTAAGAAATATTTTATTATTAATTAATAAATTTAAAAATAATTTTAAAAAAATTCAA 1663

QY 1479 atatttcaaaaatgacaacttcatcgaaaaaaatttagttatgaaaaacaactccagaatta 1538  
Db 1664 TTATATTTTATATATATATATATATATATATATTAATTTTCAATTAATTAATTAATAATA 1723  
QY 1539 aacatgactatataaatttaagataaaaatgaaatgaatattgatatglatgtaataacata 1598  
Db 1724 TAATTAATAATTAATTTTATTTAATCACTAAATCTGAAATTAATTAATTAATATATATATA 1783  
QY 1599 tctgattgcygtaaaaaaaacatactgattaaatgctcatgcaagcccaatgtaacta 1658  
Db 1784 TATATAAAAAAATGAATAATAATTTATTCCTCATTAATTAATTTATTTGTATTAATTAATA 1843  
QY 1659 tgaatgcatacagtttcttatttccaaa-----taactaatatataatccaanaata 1711  
Db 1844 ACTTAAAAAATATTTTTTTAAAAAAAATGATTTATTAATAATTAATTAATTAATTAATTA 1903  
QY 1712 gtttgcagattaaatttttctggtgctcagcttcccaacttaactaaactagtttg 1771  
Db 1904 TTTTATTAATTAATAATTTATTTTATTAATTAATAATTTATTAATTAATAATAATAATTTT 1963  
QY 1772 gaatgtctctcttatttcttcttcttctgatttcttaagtttcttaattatgaaat 1831  
Db 1964 AATATAATTAATAAATTTTAAAAATGATTTTATTAATAAATTCATATATATTTAT---ATA 2020  
QY 1832 ttaagacgagatcgttagtgctctctcttctgttcttcttaagttactcttgtaa 1891  
Db 2021 TATATACATTAATTAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTTATTTT 2080  
QY 1892 actcatctctcccaatttagacagctcaatcatatagttatcttcaatataatgctca--g 1949  
Db 2081 AATCACTAAATCTGAATTAATTAATTTGTATATATATATATATATATATATATATTAATA 2140  
QY 1950 ttgataaaaaaaatgaaaaaactcggcqqtagtcttactlaatglttlygtlaaaaaactc 2009  
Db 2141 AAAATAAATTTATTTCCCCCTATTCATAAATTTATTAATTAATAAATTAATAAATAAT 2200  
QY 2010 gatattatgaaactcaatcaatttcttctgctgataaatgctgggtlaaatttagtaattc 2069  
Db 2201 TTTTAAAAAAAATAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2259  
QY 2070 ttacataaataagaactgtaatgctgattatatttggggaaatccagtatatattagcttgg 2129  
Db 2260 TTATTTTAATAATAAATTAATTTTAAATAATTAATAAATAATTAATTAATTAATTAATTAAT 2319  
QY 2130 aactatacttcttggaata--cttgaagatttaactatttgcgaataattatagctcc 2187  
Db 2320 ATATATTAATAATGAATAATTAATTTTAAATTTTAAATAAATAATTAATTAATTAATAA 2379  
QY 2188 cgaataatacagaacgagcgacacacatalaagcaggttgaacttgcgaataatlc 2247  
Db 2380 TTAATAATAAATCTAATTCATTAATAATTAATTAATAATAATAATAATAATTAATTAATTAAT 2439  
QY 2248 tglatacaataacctataataactaattgctggttgggtccaaagcccttcaaaag 2307  
Db 2440 AATTAATAATAAATTTATTTATTTACIAAATTAATTAATTAATAATAATAATAATTAATTT 2499  
QY 2308 tcccaattt---aaattcaagacatttacaataggaataagttgagtcataaaaa 2364  
Db 2500 TTTTTTTAAATTAATAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2559  
QY 2365 taatggttatttctglaaggttttcttcttgatnaa--acgcacataatgaagaattagt 2422  
Db 2560 TATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2619  
QY 2423 ttttctcactcaaatatacaatttaatttaaaaccatgcaaccatlcataaacaataac 2482  
Db 2620 TTTTAAAAAATTTCTTAATGTAATTAATTTTATAAAAAATATTTATATAAATAATCATG 2679  
QY 2483 tattaaagaatataaataatcacaaaatattaaatacacttlaaatttcatataaattc 2542  
Db 2680 TTTTAAAAAATAAACAATAAATTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTA 2739



QY 2543 acaaacatcctaattgaacacgaaaggaagtaaaatatatacataaatgagac 2602  
 Db 2740 TTTATTTTTCATTTTAAAAAAATTTTAAAAAAATTAATTTTAAAAAA 2799  
 QY 2603 atatacctataaaaaaaagagcatatgaagtaataataagagacatgataa 2662  
 Db 2800 AC-TATATACTAATTATAAATAGATATTTATATATATAAATATTAAATATA 2858  
 QY 2663 gcattcggttaattaatcgagtcgaagatatata:cgagtaaatacatatgtgatatc 2722  
 Db 2859 TTATATATCTAATAATTAATAAATAAATTTAAATTTAAATGTAGATATAATTAT 2918  
 QY 2723 tggaaaaagaatatatatatgagaataaaga 2756  
 Db 2919 AAAAATTTATATCTCATATTTATTATTATA 2952

RESULT 13

DMU37541

LOCUS DMU37541 19517 bp DNA circular INV 02-MAR-2001  
 DEFINITION Drosophila melanogaster complete mitochondrial genome.  
 ACCESSION U37541  
 VERSION U37541.1 GI:1166529

KEYWORDS  
 SOURCE  
 ORGANISM

fruit fly.  
 Mitochondrion Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 12511 to 12682)  
 Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and  
 Wolstenholme,D.R.

TITLE Drosophila mitochondrial DNA: a novel gene order  
 JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)  
 REFERENCE 83090428  
 MEDLINE 2 (bases 5269 to 5695)  
 AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.

TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5'  
 flanking sequences and comparisons to mammalian mitochondrial tRNA  
 genes  
 JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)  
 REFERENCE 83220794  
 MEDLINE 3 (bases 404 to 5272)  
 AUTHORS de Bruijn,M.H.

TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and  
 genetic code  
 JOURNAL Nature 304 (5923), 234-241 (1983)  
 REFERENCE 83245048  
 MEDLINE 4 (bases 804 to 1778)  
 AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.

TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in  
 Drosophila melanogaster and its sibling species  
 JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)  
 REFERENCE 86174373  
 MEDLINE 5 (bases 5268 to 13619)  
 AUTHORS Garesse,R.

TITLE Drosophila melanogaster mitochondrial DNA: gene organization and  
 evolutionary considerations  
 JOURNAL Genetics 118 (4), 649-663 (1988)  
 REFERENCE 88212147  
 MEDLINE 6 (bases 441 to 2967)  
 AUTHORS Satta,Y. and Takahata,N.

TITLE Evolution of Drosophila mitochondrial DNA and the history of the  
 melanogaster subgroup  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)  
 REFERENCE 91088557  
 MEDLINE 7 (bases 14215 to 14512)  
 AUTHORS Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and

TITLE Atkinson,P.W.  
 Evidence from 12S ribosomal RNA sequences that onychophorans are  
 modified arthropods  
 JOURNAL Science 258 (5086), 1345-1348 (1992)  
 REFERENCE 93088057  
 MEDLINE

REFERENCE 8 (bases 14917 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L., Farguhar,A.L. and Kaguni,L.S.  
 TITLE Sequence, organization, and evolution of the A+T region of  
 Drosophila melanogaster mitochondrial DNA  
 JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)  
 REFERENCE 94285822  
 MEDLINE 9 (bases 1 to 408; 13319 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
 TITLE Drosophila melanogaster mitochondrial DNA: completion of the  
 nucleotide sequence and evolutionary comparisons  
 JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)  
 REFERENCE 96423163  
 MEDLINE 10 (bases 1 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,  
 Michigan State University, East Lansing, MI 48824-1319, USA

FEATURES

source

1..19517  
 /organism="Drosophila melanogaster"  
 /organelle="mitochondrion"  
 /db\_xref="taxon:7227"  
 /note="derived from new and previously submitted  
 sequences; sequence is a composite containing sequences  
 obtained from different Drosophila melanogaster strains  
 1..65  
 /product="trna-ile"  
 complement(97..165)  
 /product="trna-gln"  
 171..239  
 /product="trna-phe"  
 240..1265  
 /codon\_start=1  
 /transl\_table=5  
 /product="NADH dehydrogenase subunit 2"  
 /protein\_id="AAC47811.1"  
 /db\_xref="GI:1166530"  
 /translation="MENSSKILFITIMIGTLITVTSNSWLGAWMGLEINLSFJ  
 LSDNNLMSTEASLUKFLQVLASTVLSILMLKNNNNEINSEFTSMIMSJ  
 LKSGAPFHFWMMEGLMTWOKIAPMLISYLNKIKILLISVILSV  
 GAIGLNQTSLRKLMAFSSINHLGWLSSLMSESILFLFFYSFSEVJTFEN  
 KLFHLNQLFSWFEVNSKILKFTLFMNFSLGGLPPLGFLPKWLVIQQLTLCNQYFV  
 IMMSTLITLFFYLRLCYSAFMNMFENNIMKMNNSINYNMYMIMTFISFGLF  
 SLEFVFMF"

1264..1329  
 /product="trna-Trp"  
 complement(1322..1383)  
 /product="trna-Cys"  
 complement(1403..1468)  
 /product="trna-Tyr"  
 join(1470..1472,1474..3009)  
 /note="mechanism underlying reading frame shift after  
 first codon uncertain"

/codon\_start=1  
 /transl\_table=5  
 /product="cytochrome c oxidase subunit I"  
 /protein\_id="AAC47812.2"  
 /db\_xref="GI:7412849"  
 /translation="MSRWLFSTNKKDGTLYFTFGAWAGWGTSLILRAELGH  
 ALIGDDQIVNVITVAHAFIMFFMVMPIMGFGNMLVPLMGAPMAFPRMNMSS  
 LLPALSLIVSSMVENGASGTGWTPPIASGIAHGASVDLAIFSLLAGISSIT  
 VNFITIVINMRSTGISLDRMPIFWWSVYITALLSLPLAGATITMLIDRLNLT  
 FDPAGGDPILYQHLFWFVGHPEVYILLPFGMISHIISOESKKEFGSLGIY  
 LAIGLGFIVAAHMFIVGMDVDITRAYFTSAITIAVPTGIKIFSMATLHGTOIS  
 PAIIMALGEVFEFTVGGI.TGVVLANSSVDILHDTYVVAHFHYVLSMCAVFAIMA  
 IHWPPLFTGTLNNKMKLSHFTIMFTIGVNLTFEPQHFGLAGMPRYSVDYATYI  
 IVSTIGSTISLGLIFHFFFIWESLSQOVZYPIQLNSSIEWYONTPPAEHSYSE  
 LITN"  
 3012..3077  
 /product="trna-ileu"  
 3083..3767  
 /note="TAA stop codon is completed by the addition of 3'  
 residues to the mRNA"





QY	1119	agtaaccatgatcttgcaccacacactattctctcatcttctgtgtctatagatcttg	1178
Db	16230	AAATTCAAATTCATATATTTATATATATATACATATAAATTAATTTCAATTAAATATATAT	16289
QY	1179	cttataatgtcttctgtatctgtcgtacgtacgtatagaatttaatggttaaac	1238
Db	16290	AAGTATATAATAATATTTAT- - - - -TTAATCACTAAATCTGAATTAATTAAATGGTATA	16344
QY	1239	aaggtctataataaactagcaccaaatgagtcacatgaaatctgttagcgaaaaagtagaaa	1298
Db	16345	TATATATATATATATATATATGTAAATGAAAAATAAATTAATTTCCCTATTCATATAAT	16404
QY	1299	tatatgagctcttaaacgcataatataaatttctgtcgaaaacttagcttaacat	1358
Db	16405	TATTATATAATTAACCTTAACCAAAATATTTTAAACCAAAATTAATTTAATAAT	16464
QY	1359	gactcaatgatattctcttaattctcttgccaatlaacacatgcacgacttg	1418
Db	16465	ATACCTAATTAACCTATTTTATAATTAATTTTATTAATTAAT- - - - -TATTTAAA	16519
QY	1419	caagttatgtgcgaattcttccattccacacacactaaacttaattagaactctagaa	1478
Db	16520	TAATTAATAGCAATATTTTATATATAAATAAATAAATAATTTTAAACCAATTCAA	16579
QY	1479	atattttaaactgacacacttatacgaaaaaatttagttatgaaaaaacattccagaatta	1538
Db	16580	TTTATATTATATATATATATATATATATATTAATTTTCAATTAATTAATTAATA	16639
QY	1539	aacatgagctatataaatttaagataaatagaagtaatatgtgatatgtatgataacata	1598
Db	16640	TAATTAATAATTTATTTTAATCAGCTAAATCTGAATTAATTAATTAATATATATATA	16699
QY	1599	tctgattcggtgtaaaaaaacaacatcgttaataatgttcacgagcccatgcaacta	1658
Db	16700	TATATAAAAAATGAAAAATAAATTTATTCCTTATTCATATAAATTTATGTATAATTAA	16759
QY	1659	tgatgtcacacagtttcttatttcacaa- - - - -taactaatatataatccaataata	1711
Db	16760	ACTTAATAAATATTTTAAAAAATAATGATTTTAAATTAATTAATTAATTAATTAACIA	16819
QY	1712	gtttgtcagattaaatttcttctgtgtcagcttctccaactcaactaaactagttg	1771
Db	16820	TTTTTATTAATTAATTTATTTTAAATTAATAATTTTAAATTAATTAATAAATATTTT	16879
QY	1772	gaatgtctctcttatttcttcttcttctgtattctctatgtttttatttatgaaat	1831
Db	16880	AATATATATAAATAATTAATAATGATTTTAAATAAATAATTCATTCATATATTTAT- - -ATA	16936
QY	1832	ttaagacggatggttagtgcgttctctcttcttctgttcttcaaaagttacttctgtaa	1891
Db	16937	TATATACATATTAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTTT	16996
QY	1892	actcacctctcccaacttagacagtcacatcatagttatccttcaatatatgtc- -g	1949
Db	16997	AATCACTAAATCTGAATTAATTAATGTATATATATATATATATATATATATATATGTAATAATG	17056
QY	1950	ttgataaaaaaaatgaaaaaataactgtgtgagttctactaatgttctgtaaaaaact	2009
Db	17057	AAAAATAATTTATTCCTCATATCAATTAATTTATATTAATTAATAAATACTTAAAAAATAT	17116
QY	2010	gatatatgaatcgaatcaattctcttgatcgataaatgtgtggttaattctagtaatt	2059
Db	17117	TTTTTTTTAAAAAATAATTTATTTAATAATTAATTAATTAATAAAGCTATTTTA- -TAATAAA	17175
QY	2070	ttacataaataagaactgtaattgttgatgtatattctggggaatcagtatattagcttggt	2129
Db	17176	TTATTTTATAAATAATTAATTTAATAATTAATTAATAAATAATTTATATATATATATATAT	17235
QY	2130	aactataactctgtaata- - -ctgaagatttaactattcgcaaaatataaatttagtcc	2187
Db	17236	ATATATTTAAAAATGAATAATTTTAAATTTTAATTAATTAATTAATTAATTAATTAATTA	17295
QY	2188	cgaaaaatacagacgcgggacacgcacaacataaagcaggttgaactcttggaataatt	2247

DG	17296	TAAATAAATCGATTCTCATTTATTAATATTTTAATTAAATTAATAATTAACTAAT	17355
OY	2248	tgtatacataacctatataaalactaatgctcgttgqgtccaaaagcctttccaagg	2307
DB	17356	AATTAAATAAATATTATTATTACTAATA-TTAAATTAAATAAAAAA-TATTTTITTT	17415
OY	2308	ttccatttctt--aaattcaagacalllacatagaataaagltgagtcataaaaa	2364
DB	17416	TTTTTTTAAATAAATTAAATTAAATTATTATATATTATATTAATAATTATATTGAA	17475
OY	2365	taatggtcatlltqlaaggtlltlltlqatlcaa--accgcacatatlaaqaagtlagt	2422
DB	17476	TATTTATTAATATATATATATATAGAAAAATTAAATT-TTAAATTAATTAAATTAAT	17535
OY	2423	ttlcttccactaccaataatcaattlaattlaaaaaccatgcaccatcttcataaaaacaac	2482
DB	17536	TTTTTAAAAATTCTTAAATGTATTATTTTPATAAAAAATATTATTAATAAATCATG	17595
OY	2483	tatlaagaatatataaalcacadaatatlaaalacaccluaaattcacatataaatc	2542
DB	17596	TTTTTAAAAAATAAACAAAAAATTTTAAATAAATAAATTTTATATGAATAITAATTTA	17655
OY	2543	acaaaacatctaattlaaltgaaacagaaacgaaaggtaaaatalatacataaaalgaagac	2602
DB	17656	TTTATTTTTCATTTTTAAAAAAAATTTTTTAAAAAAAATAATTTTTTTAAAAAAA	17715
OY	2603	atatatocctataaaaaaaaaatgggcataltgaaglaaataatnagagacatgcattga	2662
DB	17716	AC-TATACTACTAATTATAAATTAATAGATATTATATATATATATTAATTTTAATATTA	17774
OY	2663	gcattcggttaaattcaatcgagtcacaaagatatataatcagtaaatcacatatgttatattc	2722
DB	17775	TTATAATCTAATAATTTAAATAAAAAATTTTAAAAATTTTAAAAATGTACATATAATTTAT	17834
OY	2723	tggaaaaagaatatatatatcgaagaaataagaaa	2756
DB	17835	AAAAATTATAITCTCATAATTTAITTATTATTA	17868

RESULT	14
LOCUS	ABO20742
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K21H1.
ACCSSION	ABO20742 BA000015
VERSION	ABO20742.1 GI:3985931
KEYWORDS	
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC clone:K21H1.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyt Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi. 1 (sites)
REFERENCE	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tabata,S. Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixt P1 and TAC clones DNA Res. 7 (1), 31-63 (2000) 20181125
JOURNAL MEDLINE	2 (bases 1 to 74342) Nakamura,Y. Direct Submission Submitted (03-DIC-1998). Yasukazu Nakamura, Kazusa DNA Research institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp, Tel.:81-438-52-3935, Fax:81-438-52-3934)
AUTHORS	TITLE:
JOURNAL	COMMENT
	For the latest information on annotation of this clone, please s <a href="http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K21H1">http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K21H1</a> Genes with similarity to proteins in the databases are described 'product' or 'note' qualifiers. Genes that have no significant

protein similarity are described as 'unknown protein'.  
The software programs used to predict genes include: Grail  
(Informatics Group, Oak Ridge National Laboratory,  
http://compbio.ornl.gov/Grail-1.3/),  
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),  
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of  
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and  
SplicePredictor (Volker Brendel, Stanford University,  
http://gremli1.zool.iastate.edu/cgi-bin/sp.cgi).  
Genes encoding tRNAs are predicted by tRNAscan-SE  
(Sean Eddy, Washington University School of Medicine, St. Louis,  
http://genome.wustl.edu/eddy/tRNAscan-SE/).  
This sequence may not be the entire insert of this clone. It may be  
shorter because we remove overlaps between neighboring submissions.  
The 5' clone is K8A10 and the 3' clone is K3G17.

FEATURES  
source  
Location/Qualifiers  
1. 74342

/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="K21H1"  
/clone\_lib="Mitsui TAC"  
complement(join(683..778,1044..1247,1325..1959,2053..2283,  
2364..2661))  
/note="gene\_id:K21H1.1  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB10939.1"  
/db\_xref="GI:10177592"  
/translation="MMNSDDDDDEBRGYLLRPEELRPWELVRLLESFSDIEKPRSDSS  
ETEEHSFRHRWLI FVSLVLLKLLRFFSKLLALVGSALFSENLFSNNSFSGFLRGEV  
VMPQRTSENYQSFIGHLDTRVSLDMLTNRDEGEKYAALSIMASKIAYENARIKHV  
ENHWNMRYLGLVDYWNEYQEKETQAIFMSTDETTTRNGQETTVVAFRGTELFNSE  
DMCSDFDITWFLPNIGNIHGSPKALGLQNNCSMPKEPLSNPDRKSPLAYYSTRDSL  
KTLIAQNKNTKFLVTLGSLGALALFTAVLIHETELLERIQGVYTGQPRVQDSK  
FGEFMEKLEKYNIKYRFVYNNDIVRPLPYDDKDLMEKHFCTCIYDQNTQAKVME  
QSDENFLLRGIIKMMMSAILEFIRSFIVAEGSEYSEGWLLKGRALGIIVGVSN  
HTPDYKNSRVVSINGYEDVPRKDETALKRVIFIPFG"  
7772..8497

CDS

/note="emb|CAB62312.1  
gene\_id:K21H1.2  
similar to unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB10940.1"  
/db\_xref="GI:10177593"  
/translation="MDSDIMNMNMHOMKLEPFCNPNSFFSPDHNTYPLFNSTHY  
OSDHSMTNEPFRYSGGLTNPSSISPTAYSSVFLDKRNNNNNNNGTMMAMREMI  
FRIAYMODIHIDPEAVKPPKRRARYISKDPSVAAHRHRRISERIRILORLVPGGTK  
MDTASMLDEAIIHYVKFLKKQVSLSEQAVVTGGGGGGGRVLIGGGMTAASGGGGGG  
GVVMKGGTGVGTQHMVGNQILR"  
9616..10005  
/note="gb|AAF64534.1  
gene\_id:K21H1.3  
similar to unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB10941.1"  
/db\_xref="GI:10177594"  
/translation="MAASSLNLTLILSLTFLISLQSESLSDNPSLTLLPDGFDPIS  
HSDEFDIDGEESFEVTEEDDGYDRSLYWRRTKYIISYGALSANRVPCPPRSGRSY  
YTHNCFRAGPVPHPYSRGCSSTRCRR"  
14222..15256

CDS

/note="gene\_id:K21H1.4"  
/codon\_start=1  
/evidence=not\_experimental  
/product="protein kinase-like protein"  
/protein\_id="BAB10942.1"  
/db\_xref="GI:10177595"  
/translation="MEMIRGETIGYGTSTVSLATRSNNDGCEFPPLMAVKADS YGA  
ASLANEKSVLDNLGDDCNEIVRCFGEDRTVENGEEMHNLLEYASRGSLESYLKLAG

CDS

CDS

EGVESTVRRHTGSVLRGLRH:HANGFAMCDLKIQLNILEGDAVKIADFGLAKRIG  
LTALNKGQIRGTPLYMAPESVNDNDEYGESEGDVWALGVVYEMFSKGTAMSLKEGSH  
MSLLRLRIGDEVPMIPEELSEQGRDFLSKCFVKDPPKRRTAEMLLNHPFVTVDDH  
VLVKEEDFVVMNMTEDVSTSPRCFFEPDPWVSVSSGSQITDSPDRVASLVTDMLPC  
SVINSWTVVR"  
complement(15607..17817)  
/note="gene\_id:K21H1.5"  
/codon\_start=1  
/evidence=not\_experimental  
/product="subtilisin-type protease-like"  
/protein\_id="BAB10943.1"  
/db\_xref="GI:10177596"  
/translation="MGMTVILIVSEFFVAIVTAETSPYITHMDLSAKPLPFSQHR  
FSTLTISVITNRKPKIYAYTDSVHGSFSAVLINSELQRLKHPGVSFTKDLPVKLH  
TFSPKFILGNSTSGTWPVSNYAGAGIYIGIIDTIGIMPDSPSFHDGVSVPKMKKAC  
FNSSSLCNKKLIGAKVFENKGLFANNPDLRETKIGQYSSPYDTIGHGTHVAIAAGNH  
KNASYFSTAOGTASGIAPHALAIYKAAMEGITYSSDVIAIDQAIRDGVHVISLSL  
LSFEDDDDDNGFLENDPIAVASFAIQGVFVITSGNDGPPYWSLINGAPWIMTV  
AGTIGRFQGITLTFGNRVSFSPSLRPGEPSPVOPVTYIEGSEVENKTLANRIYVC  
ENINIGSKLHQIRSTGAADVLLIDKLEEDTIFKQFPVAFVIGSKHRETIEMVASS  
KNMATAKLEFRKTVIGTKPAPEVGTYSRGPFTSPFOILKPDILADGTILSAMPV  
OITGTRALPLFSGENLLTGTSMAPHVAGVAAALIKOVHPMSBSAIFKSAIMTALT  
NPLAVGAGVSTNKNVLPGLIYDTTPROFINFLCHFAKQSKRLINIITRSNISACK  
PSPYLNPSTIIAYFITSDDSPKIFKRLITNVGEAKRSYIVRGLKGLNVVEPKKI  
FSEKNEKISTVRLESFPRGLQENYVYGLVSWDEDEAEFEVSCSVATSLVQES"  
join(18490..19596,19689..19859,19945..20028,20117..20  
20278..20342,20515..20605,20698..20883,21116..21279,  
21388..21502,21780..21956,22033..22146,22391..22510,  
22590..22904,22985..23074,23461..23579,23668..23740,  
23831..23887,23999..24091,24221..24343,24422..24519,  
24598..24643,24730..24843,24949..25032,25116..25322,  
25414..25490,25698..25821,25946..25999,26108..26227,  
26298..26372,26460..26600)  
/note="gene\_id:K21H1.6"  
/codon\_start=1  
/evidence=not\_experimental  
/product="DNA polymerase alpha 1"  
/protein\_id="BAB10944.1"  
/db\_xref="GI:10177597"  
/translation="MSGDNSTETGRRHSRGAEASSRKQTLERKAIROGGIRASAGG  
YDIRLOKPIFDIVDDEYDALVSRREEARGFVEVDEGEGDGLYIDEGEEDWMSKPS  
PESTIDESDGRFSGRLKKKKKKKEQTOQPQVKKVNPALKAAATITGEGRLSMFTS  
SFKKVKETDKAQYEGILDEIIAQVTPDESDDKKHTRKLPQVPTLFFNKKLFSVA  
SMGKSESEPTPSTYEGDSVSMDELKEEDEKSEVTPSEIMELGSDIVKEDGSNK  
RKTEVKSGLGVKEVFTLNATIQMKEDSALATAGWKEAMGKVGTENGAILGSSSFG  
TEFDLDADGSLRFFILDAYEEAFGASNGTYLFGKLVKMGDIYKSCVVKNIQRCVY  
IPNDSIFPSHELLIMLEQEVKDSRLSPESFRGLHEMASKLKNELAQEJLQNVNSF  
APVKRNVAFERPDVPAGEQYVLKINYSFKDRPLPEDLKGESFALLGSHTSALEHFI  
KRKIMGPCMLKISSFSTCSPSEGVSMCKFEVTVOSPKDITIVYSEKVVHPRAVITA  
NKTIVNEKQNISEIVSASVLCFNNAKIDVPPAPAPERKRSGLISHTVVRNPEGIGY  
IGKKEVSDRNSKNCVNLSENSEALLNRLFLFLNKLKDSIDLIVGHNISGFDLVL  
QRAQACKYQSSMSKIGRLKRSFMPKLGKNSNYGSGATPGLSCTAGRLLCDTDLCS  
DLKEVSYSLDLSKTQLNDRKELAPNDIPKMFQSSKTLVELLEGETDAMLSEML  
FHLVLPPLTLQLTNISGNLMWKTLOGARAQRIEYVLLHFFSKFFILPDKISQRMKE  
KSSKRRMAYABEDRNVDLADLTLLENDPSKSKTKKGPAYAGGLVLEPKRGLYDKY  
LLIDFNSLYPSIIIOEYNICFTTIPRSEDGVPRLPSSQIPGILPKLMEHLVSIRKSVK  
KMKKETGLKYLWELDIRQALKLTANSMYGCLGFSNSRFTAKPLAELITLQGRDILQK  
VDLYONHNLNDEVITGDTDSIMHSGIDDIIEVKAIKSKYIOEVNKKYRCLKIDCDGI  
KRMILLRKKKYAAAVKLQFKDGRKCEBDEKRGVDMVRBMSLSKEIGDLCLSKILYG  
SCEDVVEAIIHNEMLMKIKEEMRNGVALEKYVITKTLTKPPAAYPDSKSOPHVOVALPR  
RQRGYKEGFNAKDTPYIICYEGGNASSASSAGIAERARHDEVKSEGSRLVIDIDY  
LAQOIHVAVSRCAEIQGTSPEPLAECIGLDPKYSRYSKNDATSSDPSILFATSD  
ESKKPATPETEESDSTFWLKLHCPCQOEDSTGIIISPAMIANOVKROIDGFVSMYK  
IMVCEDESCKHTRTSPNFRILGERGTGCPNYPNCGNFTLRKYTEADLYKOLSYFC  
ILDTQCSLEKMDVGVRIOVEKAMTKIRPAVKSAAAITRSSRDRCAIGWQTLDIVI"  
complement(join(26828..27019,27114..27179,27258..27323  
27414..27485,27598..27834))  
/note="emb|CAB16798.1  
gene\_id:K21H1.7  
similar to unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB10945.1"

CDS

```
/db_xref="GI:10177598"
/translation="MGDSVDVGRLPSPSSDEISFLRQILSRTPTAPSSPPKSTNV
SSAEFFPSVSGAVSVGVSETGQDKYAFERKRSKAKORNSLKRNI DAOFHNLS
KKRSKINEMKALQKLI PNSNKTDKASMLDEAIEYLKQLQOVOT-AVMNGLNPM
RLPQVPPPTHTRINETLEODLNL ETLLAAPHSTLEPAKTSQGMCFSTATLL"
complement(join(29326..29586,29960..30082,30390..30512,
30845..30958,31036..31233))
/note="gene_id:K21H1.8
pir|T05432
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10946.1"
/db_xref="GI:10177599"
/translation="MDPNEDNSEIPDSFSPFWLLFGDDMDIMDTYSPVDADISPDF
TLVDNGDPAIEPGLRITISMETTEQDSLOSWMNDEQSETTSVEYTDVSSHGNFT
NEEETLERWYRNWLOSSINEQSETSGQEYTNASSHGGTFIYEETLEQYWRNWLQSS
TNEOFETESLEEYTNPSHGDI FYEELSLTDETDERTGLSEVIDENLIRKYEK
RSDDETRKVCVICOQKLDNEEVSKLGCGHDPHFGCIKNWLMVTNKCPLCNREVV"
join(33235..33402,33654..33791,33873..33971,34064..34367,
34837..35225)
/note="contains similarity to MAP3K-like protein kinase
gene_id:K21H1.9"

Query Match      3.2%; Score 125.2; DB 8; Length 74342;
Best Local Similarity 64.5%; Pred. No. 3.2e-09;
Matches 187; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 3015 acatgatacgatgaagaagatgcagtcacatgcgcgcacgacccgtagacatcg 3074
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8052 ACATGAGAGCTATGCGAGAGATGATCTTCGATCGCCGCGTATGCAACCGATCCATATCG 8111

QY 3075 accctgcacggtccctaagccgaaccgcgtaacgtaagataagcgacgacccacaga 3134
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8112 ATCCCGAGGCGGTTAAGCCACCGAAGAGAGAAACGTCAGATCTCTAAAGATCCTCAA 8171

QY 3135 cgtgtgtgtctcgtcgcgcgcggaagaagatcagcgagaagaatccgaatctcaagaagaa 3194
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8172 GCGTGGCGGCTAGGCATAGAAAGGAGAGAAATAAGCGAGAGAGATTCGGATTTTGCAACGGC 8231

QY 3195 tcgtgcctgtgtgtcgaagaatgaacacagcttcacatgcgcgaagaagccatacgttaca 3254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8232 TTGTTCTGTGCTGGAGCAGAAAGATGATACAGCTTCGATGCTCGATGAAGCAATTCAATTATG 8291

QY 3255 ccaagtccttgaacgcgacgtagaagatcttcacgacctcacttcagatt 3304
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8292 TGAAGTTTITAAAGAAACAGGTGCGAGTCTCTGGAGGAGCAGCGCGTGGTT 8341

RESULT 15
LOCUS AC092304/c 36977 bp DNA HTG 03-JUL-2001
DEFINITION Homo sapiens chromosome 19 clone LNLPOS-20C5, WORKING DRAFT
SEQUENCE. 5 unordered pieces.
ACCESSION AC092304
VERSION AC092304.1 GI:14589493
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFLN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 36977)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 36977)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
```

```
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 31883, FOS36853
Center clone name: LNL-FOS_20C5
-----
Summary statistics
Consensus quality: 31111 bases at least Q40
Consensus quality: 33034 bases at least Q30
Consensus quality: 33871 bases at least Q20
Estimated insert size: 42980; agarose-1p estimation
Estimated insert size: 36477; sum-of-contigs estimation
Quality coverage: 6.56 in Q20 bases; agarose-1p estimation
Quality coverage: 7.73 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1580: contig of 1580 bp in length
*
* 1581 1680: gap of unknown length
*
* 1581 3748: contig of 2068 bp in length
*
* 3749 3848: gap of unknown length
*
* 3849 6070: contig of 2222 bp in length
*
* 6071 6170: gap of unknown length
*
* 6171 13635: contig of 7465 bp in length
*
* 13636 13735: gap of unknown length
*
* 13736 22951: contig of 9216 bp in length
*
* 22952 23051: gap of unknown length
*
* 23052 36977: contig of 13926 bp in length.

FEATURES
Source
Location/Qualifiers
1..36977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LNLPOS-20C5"

BASE COUNT 11122 a 7407 c 7254 g 10688 t 506 others
ORIGIN

Query Match      3.2%; Score 124.8; DB 2; Length 36977;
Best Local Similarity 45.3%; Pred. No. 3.8e-09;
Matches 745; Conservative 0; Mismatches 873; Indels 27; Gaps

QY 1165 tctataggatttgccttataatgttcttcgtatgcgcgaagcagtaagtaagatt 1224
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34403 TATATATATATATTAATAATGATATATATATATATATATATATATATATATATATAT 34344

QY 1225 caatggtataacaagattatataaactagacaaaagatccatgaattgttagc 1284
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34343 AATTATATATATATTAATAATACATTATATATATATATATATATATATATATATATAT 34284

QY 1285 gaaaaaggtagaataatattgagcttlaaacggaatatatatataatttgcgcgaaac 1344
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34283 GTAATAATACATTAATATCTATATAATTTATTTATATATATATATATATATATATATAA 34224

QY 1345 ttagcttaatcatgactaaatgatatatttccttaatttccttgcgaattacacat 1404
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34223 TTATATTAATTAATTAATTTATATGTTAATAATTAATTTTATATATATATATATATAT 34164

QY 1405 gcaacgattttgtgcaagttatgtgtcgaattctccattccacacaacactaaactaac 1464
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34163 ATGTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 34104

QY 1465 tagaactctagaaatatlttaaatgacaacttatacgaaaadaalltagttatgaaa 1524
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34103 AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 34044

QY 1525 caattccagaattaaacatgagctatataaatttaagataaaatgaagtaacattgatatg 1584
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```





